PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A2

(11) International Publication Number: WO 00/58516

(43) International Publication Date: 5 October 2000 (05.10.00)

(21) International Application Number: PCT/US00/08069

(22) International Filing Date: 27 March 2000 (27.03.00)

(30) Priority Data:

60/126,473	26 March 1999 (26.03.99)	US
60/140,359	23 June 1999 (23.06.99)	US

(71) Applicants: WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).

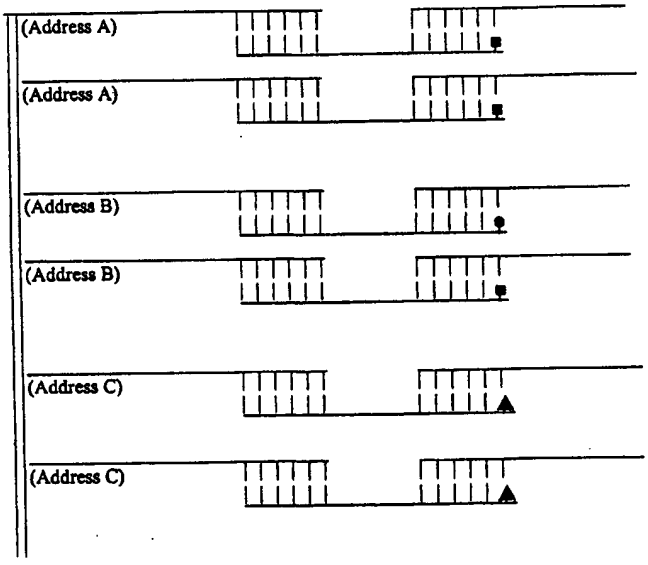
(72) Inventors: FAN, Jian-Bing; 3737 Nobel Drive, #2115, San Diego, CA 92122 (US). HIRSCHHORN, Joel, N.; 14 Oakmont Road, Newton, MA 02459 (US). HUANG, Xiaohua; 937 Jackson Street, Mountain View, CA 94043 (US). KAPLAN, Paul; 2190 Vizcaya Circle, Campbell, CA 95008 (US). LANDER, Eric, S.; 151 Bishop Allan Drive, Cambridge, MA 02139 (US). LOCKHART, David, J.; 510 Torrey Point Road, Del Mar, CA 92014 (US). RYDER, Thomas; 100 Spring Street, Los Gatos, CA 95030 (US). SKLAR, Pamela; 68 Fuller Street, Brookline, MA 02446 (US).

(74) Agents: HOGLE, Doreen, M. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: UNIVERSAL ARRAYS



(57) Abstract

An array of oligonucleotides on a solid substrate is disclosed, which can be used for multiple purposes. Methods and reagents are provided for performing genotyping to determine the identity or ration of allelic forms of a gene in a sample. A single base extension primer is coupled to a sequence identity code. During the primer extension reaction a distinctive label is incorporated which identifies the allelic form present in the sample. This permits multiple simultaneous analyses to be performed easily and efficiently.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

UNIVERSAL ARRAYS

BACKGROUND OF THE INVENTION

Obtaining genotype information on thousands of polymorphic markers in a highly parallel fashion is becoming an increasingly important task in mapping disease loci, in identifying quantitative trait loci, in diagnosing tumor loss of heterozygosity, and in performing linkage studies. A currently available method for simultaneously obtaining large numbers of polymorphic marker genotypes involves hybridization to allele specific probes on high density oligonucleotide arrays. In order to practice the method, redundant sets of hybridization probes, typically twenty or more, are used to score each marker. A high degree of redundancy is required, however, to reduce the noise and achieve an acceptable level of accuracy. Even this level of redundancy is often insufficient to unambiguously score heterozygotes or to quantitatively determine allele frequency in a population. Thus, there is a need in the art for more reliable and better quantitative methods to identify genotypes at polymorphic markers.

SUMMARY OF THE INVENTION

An array of oligonucleotide tags attached to a solid substrate is disclosed, along with locus-specific tagged oligonucleotides. The array and the locus-specific tagged oligonucleotides are particularly useful in genotyping using single base extension reactions. When used together, the array and the locus-specific tagged oligonucleotides serve as a "universal chip" system for use in genotyping, wherein by using different sets of locus-specific tagged oligonucleotides the system can be tailored to any desired genotyping application. For example, it is an object of the present invention to provide a method to aid in determining a ratio of alleles at a polymorphic locus. It is another object of the invention to provide a set of primers for use in determining a ratio of nucleotides present at a polymorphic locus.

Thus, in one embodiment the invention relates to an array comprising one or more oligonucleotide tags fixed to a solid substrate, wherein each oligonucleotide tag comprises a unique known arbitrary nucleotide sequence of sufficient length to hybridize to a locus-specific tagged oligonucleotide, wherein the locus-specific tagged oligonucleotide has at its first end nucleotide sequence which hybridizes to, e.g., is complementary to, the arbitrary sequence of the oligonucleotide tag, and wherein the locus-specific tagged oligonucleotide has at a second end nucleotide sequence complementary to target polynucleotide sequence in a sample.

In one embodiment, the invention relates to a kit comprising an array comprising one or more oligonucleotide tags fixed to a solid substrate, wherein each oligonucleotide tag comprises a unique known arbitrary nucleotide sequence of sufficient length to hybridize to a locus-specific tagged oligonucleotide, and one or more locus-specific tagged oligonucleotides, wherein each locus-specific tagged oligonucleotide has at its first (5') end nucleotide sequence which hybridizes to, e.g., is complementary to, the arbitrary sequence of a corresponding oligonucleotide tag on the array, and has at it's second (3') end nucleotide sequence complementary to target polynucleotide sequence in a sample.

The invention further relates to a method of genotyping a nucleic acid sample at one or more loci, comprising the steps of obtaining a nucleic acid sample to be tested; combining the nucleic acid sample with one or more locus-specific tagged oligonucleotides under conditions suitable for hybridization of the nucleic acid sample to one or more locus-specific tagged oligonucleotides, wherein each locus-specific tagged oligonucleotide comprises a nucleotide sequence capable of hybridizing to a complementary sequence in an oligonucleotide tag and a nucleotide sequence complementary to the nucleotide sequence 5' of a nucleotide to be queried in the sample, thereby creating an amplification product-locus-specific tagged oligonucleotide complex; subjecting the complex to a single base extension reaction, wherein the reaction results in the addition of a labeled ddNTP to the locus-specific tagged oligonucleotide, and wherein each type of ddNTP has a label that can be distinguished from the label of the other three types of ddNTPs; contacting the complex with an oligonucleotide array comprising one or more oligonucleotide tags fixed to a solid substrate under suitable hybridization conditions, wherein each oligonucleotide tag comprises a unique arbitrary sequence complementary and of sufficient length to hybridize to a complementarysequence in a locus-specific tagged oligonucleotide, whereby the complex hybridizes to a specific oligonucleotide tag on the array; and assaying the array to determine the labeled ddNTPs present in the complex hybridized to one or more oligonucleotide tags, thereby determining the genotype of the queried nucleotide in the sample. In one embodiment the nucleic acid sample to be tested is amplified.

In one embodiment a method is provided to aid in determining a ratio of alleles at a polymorphic locus in a sample. A pair of primers is used to amplify a region of a nucleic acid in a sample. In one embodiment, the region comprises a polymorphic locus, and an amplified nucleic acid product is formed which comprises the polymorphic locus. The amplified nucleic acid product is used as a template in a single base extension reaction with an extension primer, forming a labeled extension primer. The extension primer (also called a locus-specific tagged oligonucleotide herein)

comprises a 3' portion and a 5' portion. The 3' portion is complementary to the amplified nucleic acid product and terminates one nucleotide 5' to the polymorphic locus. The 5' portion is not complementary to the amplified nucleic acid product. A labeled dideoxynucleotide which is complementary to the polymorphic locus is coupled to the 3' end of the extension primer. Each type of dideoxynucleotide present in the reaction bears a distinct label. The 5' portion of the extension primer is hybridized to one or more probes (also called oligonucleotide tags herein) which are immobilized to known locations on a solid support. The probes comprise a nucleotide sequence which is complementary to the 5' portion of the extension primer.

Also provided by the present invention is a set of primers for use in determining a ratio of nucleotides present at a polymorphic locus. The set includes a pair of amplification primers and an extension primer. The pair of primers prime synthesis of a region of double stranded nucleic acid which comprises a polymorphic locus. The extension primer comprises a 3' portion which is complementary to a portion of the region of double stranded nucleic acid and a 5' portion which is not complementary to the region of double stranded nucleic acid. The extension primer terminates one nucleotide 5' to the polymorphic locus. Examples of primers according to the invention are shown in Table 1.

Another embodiment of the invention provides a method to aid in determining a ratio of alleles at a polymorphic locus in a sample. Any nucleic acid molecule, including genomic DNA, which comprises one or more polymorphic locus is used as a template in a single base extension reaction with an extension primer, forming a labeled extension primer. The extension primer comprises a 3' portion and a 5' portion. The 3' portion is complementary to the nucleic acid molecule and terminates one nucleotide 5' to the polymorphic locus. The 5' portion is not complementary to the nucleic acid molecule. A labeled dideoxynucleotide which is complementary to the polymorphic locus is coupled to the 3' end of the extension primer. Each type of dideoxynucleotide present in the reaction bears a distinct label. The 5' portion of the extension primer is

hybridized to one or more probes which are immobilized to known locations on a solid support.

These and other embodiments of the invention which are described in more detail below provide the art with methods and tools for rapidly and easily determining genotypes of individuals and allele frequencies in populations.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagram of the universal array. The solid substrate (*e.g.*, a glass slide) is depicted on the left, and different oligonucleotide tags ("A", "B", "C", etc.) are shown attached to the solid substrate. The nucleotide sequence on the right-hand end of each oligonucleotide tag ("Tag A", Tag B", "Tag C") is arbitrary unique sequence; that is, it is designed and synthesized to be unique to each oligonucleotide tag.

Fig. 2 is a diagram depicting a locus-specific tagged oligonucleotide. The nucleotide sequence at the left-hand end is complementary to the arbitrary sequence of one of the oligonucleotide tags depicted in Fig. 1. The nucleotide sequence at the right-hand end is complementary to the amplification product of a known polymorphic locus (*e.g.*, a single nucleotide polymorphism (SNP)). Therefore, locus-specific tagged oligonucleotide "A" comprises anucleotide sequence complementary to the arbitrary sequence of the "Tag A" oligonucleotide tag depicted in Fig. 1, and also comprises sequence complementary to SNP "A".

Fig. 3 is a diagram showing the hybridization of the locus-specific tagged oligonucleotide to the amplification product. The locus-specific sequence (right hand end) of the oligonucleotide is designed so that it terminates one nucleotide immediately before (5' of) the nucleotide to be genotyped (shown in box).

Fig. 4 is a diagram depicting the labeling of the locus-specific tagged oligonucleotide-amplification primer complex via single base extension. During the reaction, a single labeled ddNTP complementary to the queried nucleotide is enzymatically added to the 3' end of the locus-specific tagged oligonucleotide. The nucleotide is shown in the box.

Fig. 5 is a diagram depicting the hybridization of the complex of the amplification product and the locus-specific tagged oligonucleotide to the oligonucleotide tags on the array. The solid substrate to which the oligonucleotide tags of the array are bound is shown on the left, with the individual addresses labeled as "A", "B", etc. Each oligonucleotide tag is shown at its address. The locus-specific tagged oligonucleotide is shown hybridized to the oligonucleotide tag, and the amplification product is in turn bound to the locus-specific tagged oligonucleotide. The locus-specific tagged oligonucleotide is bound to a labeled (■, ●, etc.) nucleotide as a result of single base extension. Although a single complex is shown at each address, in reality, many such oligonucleotide tags are located at each address; that is, the substrate surface at address "A" has many copies of oligonucleotide tag "A" attached to it, etc.

Fig. 6 is a diagram depicting the hybridization as in Fig. 5, but the sample at address "B" is heterozygous for the queried nucleotide.

Fig. 7 is a schematic showing the combined use of amplification, single base extension of a tagged primer, and hybridization to a tag array.

Fig. 8 shows a quantitative measurement of allele frequency. Template-T (5'-TGCTGAATATTCAGATTCTCTAGTGCTACCTGAAAGATCCTG-3'; SEQ ID NO: 1) and

Template-G

(5'-TGCTGAATATTCAGATTCTCGAGTGCTACCTGAAAGATCCTG-3'; SEQ ID NO: 2) were mixed at different ratios (6 nM /60 nM, 6 nM /18 nM, 6 nM /6 nM, 18 nM /6 nM, 60 nM /6 nM, 180 nM /6 nM). Six SBE primers

(5'-CACCATGCTCACAATGAATGCAGGATCTTTCAGGTAGCACT-3' (SEQ ID NO: 3);

5'-GATAATTCTCTGATAGGCCGCAGGATCTTTCAGGTAGCACT-3' (SEQ ID NO: 4);

5'-GACTACGATGTGATCCGTGTCAGGATCTTTCAGGTAGCACT-3' (SEQ ID NO: 5);

Hybridization Conditions

The term "stringent conditions" refers to conditions under which a probe will hybridize to its target subsequence, but with only insubstantial hybridization to other sequences or to other sequences such that the difference may be identified. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

The T_m is the temperature, under defined ionic strength, pH, and nucleic acid concentration, at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. As the target sequences are generally present in excess, at T_m, 50% of the probes are occupied at equilibrium). Typically, stringent conditions will be those in which the salt concentration is at least about 0.01 to 1.0 M concentration of a Na or other salt at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides). Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide.

The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule substantially to or only to a particular nucleotide sequence or sequences under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) of DNA or RNA. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids. Under low stringency conditions (*e.g.*, low temperature and/or high salt) hybrid duplexes (*e.g.*, DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the annealed sequences are not perfectly complementary. Thus, specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency (*e.g.*, higher temperature or lower salt) successful hybridization requires fewer mismatches.

One of skill in the art will appreciate that hybridization conditions can be selected to provide any degree of stringency. In a preferred embodiment, hybridization

is performed at low stringency, in this case in 6X SSPE-T at 37°C (0.005% Triton X-100), to ensure hybridization. and then subsequent washes are performed at higher stringency (*e.g.*, 1 X SSPE-T at 37°C) to eliminate mismatched hybrid duplexes. Successive washes can be performed at increasingly higher stringency (*e.g.*, down to as low as 0.25 X SSPE-T at 37°C to 50°C) until a desired level of hybridization specificity is obtained. Stringency can also be increased by addition of agents such as formamide. Hybridization specificity can be evaluated by comparison of hybridization to the test probes with hybridization to the various controls that can be present (*e.g.*, expression level control, normalization control, mismatch controls, *etc.*).

In general, there is a tradeoff between hybridization specificity (stringency) and signal intensity. Thus, in a preferred embodiment, the wash is performed at the highest stringency that produces consistent results and that provides a signal intensity greater than approximately 10% of the background intensity. Thus, in a preferred embodiment, the hybridized array can be washed at successively higher stringency solutions and read between each wash. Analysis of the data sets thus produced will reveal a wash stringency above which the hybridization pattern is not appreciably altered and which provides adequate signal for the particular oligonucleotide probes of interest.

The stability of duplexes formed between RNAs or DNAs are generally in the order of RNA:RNA > RNA:DNA > DNA:DNA, in solution. Long probes have better duplex stability with a target, but poorer mismatch discrimination than shorter probes (mismatch discrimination refers to the measured hybridization signal ratio between a perfect match probe and a single base mismatch probe). Shorter probes (*e.g.*, 8-mers) discriminate mismatches very well. but the overall duplex stability is low.

Altering the thermal stability (T_m) of the duplex formed between the target and the probe using, *e.g.*, known oligonucleotide analogues allows for optimization of duplex stability and mismatch discrimination. One useful aspect of altering the T_m arises from the fact that adenine-thymine (A-T) duplexes have a lower T_m than guanine-cytosine (G-C) duplexes. due in part to the fact that the A-T duplexes have two hydrogen bonds per base-pair. while the G-C duplexes have three hydrogen bonds per

base pair. In heterogeneous oligonucleotide arrays in which there is a non-uniform distribution of bases, it is not generally possible to optimize hybridization for each oligonucleotide probe simultaneously. Thus, in some embodiments, it is desirable to selectively destabilize G-C duplexes and/or to increase the stability of A-T duplexes. This can be accomplished, *e.g.*, by substituting guanine residues in the probes of an array which form G-C duplexes with hypoxanthine, or by substituting adenine residues in probes which form A-T duplexes with 2,6 diaminopurine or by using tetramethyl ammonium chloride (TMACl) in place of NaCl.

Altered duplex stability conferred by using oligonucleotide analogue probes can be ascertained by following, *e.g.*, fluorescence signal intensity of oligonucleotide analogue arrays hybridized with a target oligonucleotide over time. The data allow optimization of specific hybridization conditions at, *e.g.*, room temperature.

Another way of verifying altered duplex stability is by following the signal intensity generated upon hybridization with time. Previous experiments using DNA targets and DNA chips have shown that signal intensity increases with time, and that the more stable duplexes generate higher signal intensities faster than less stable duplexes. The signals reach a plateau or "saturate" after a certain amount of time due to all of the binding sites becoming occupied. These data allow for optimization of hybridization, and determination of the best conditions at a specified temperature.

Methods of optimizing hybridization conditions are well known to those of skill in the art (*see, e.g., Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 24: Hybridization With Nucleic Acid Probes*, P. Tijssen, ed. Elsevier, N.Y., (1993)).

Signal Detection

The hybridized nucleic acids can be detected by detecting one or more labels attached to the target nucleic acids. The labels can be incorporated by any of a number of means well known to those of skill in the art. However, in a preferred embodiment, the label is incorporated by labeling the primers prior to the amplification step in the

preparation of the target nucleic acids. Thus, for example, polymerase chain reaction with labeled primers will provide a labeled amplification product.

Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical, or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (*e.g.*, Dynabeads™), fluorescent dyes (*e.g.*, fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (*e.g.*, ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (*e.g.*, horseradish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.*, polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels can be detected using photographic film or scintillation counters, fluorescent markers can be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and colorimetric labels are detected by simply visualizing the colored label. One method uses colloidal gold label that can be detected by measuring scattered light.

Means of detecting labeled target nucleic acids hybridized to the probes of the array are known to those of skill in the art. Thus, for example, where a colorimetric label is used, simple visualization of the label is sufficient. Where a radioactive labeled probe is used, detection of the radiation (*e.g.* with photographic film or a solid state detector) is sufficient.

Detection of target nucleic acids which are labeled with a fluorescent label (*i.e.*, a "color tag") can be accomplished with fluorescence microscopy. The hybridized array can be excited with a light source at the excitation wavelength of the particular fluorescent label and the resulting fluorescence at the emission wavelength is detected.

The excitation light source can be a laser appropriate for the excitation of the fluorescent label.

The confocal microscope can be automated with a computer-controlled stage to automatically scan the entire high density array, *i.e.*, to sequentially examine individual probes or adjacent groups of probes in a systematic manner until all probes have been examined. Similarly, the microscope can be equipped with a phototransducer (*e.g.*, a photomultiplier, a solid state array, a CCD camera, *etc.*) attached to an automated data acquisition system to automatically record the fluorescence signal produced by hybridization to each oligonucleotide probe on the array. Such automated systems are described at length in U.S. Patent No: 5,143,854, PCT Application 20 92/10092, and copending U.S. Application Ser. No. 08/195,889, filed on February 10, 1994. Use of laser illumination in conjunction with automated confocal microscopy for signal detection permits detection at a resolution of better than about 100 μm, more preferably better than about 50 μm, and most preferably better than about 25 μm.

Two different fluorescent labels can be used in order to distinguish two alleles at each marker examined. In such a case, the array can be scanned two times. During the first scan, the excitation and emission wavelengths are set as required to detect one of the two fluorescent labels. For the second scan, the excitation and emission wavelengths are set as required to detect the second fluorescent label. When the results from both scans are compared, the genotype identification or allele frequency can be determined.

Quantification and Determination of Genotypes

The term "quantifying" when used in the context of quantifying hybridization of a nucleic acid sequence or subsequence can refer to absolute or to relative quantification. Absolute quantification can be accomplished by inclusion of known concentration(s) of one or more target nucleic acids (*e.g.*, control nucleic acids such as Bio B, or known amounts the target nucleic acids themselves) and referencing the hybridization intensity of unknowns with the known target nucleic acids (*e.g.*, through

generation of a standard curve). Alternatively, relative quantification can be accomplished by comparison of hybridization signals between two or more genes, or between two or more treatments to quantify the changes in hybridization intensity and, by implication. the frequency of an allele. Relative quantification can also be used to merely detect the presence or absence of an allele in the target nucleic acids. In one embodiment, for example, the presence or absence of the two alleles of a marker can be determined by comparing the quantities of the first and second color tag at the known locations in the array, *i.e.*, on the solid support, which correspond to the allele-specific probes for the two alleles.

A preferred quantifying method is to use a confocal microscope and fluorescent labels. The GeneChip® system (Affymetrix, Santa Clara, CA) is particularly suitable for quantifying the hybridization; however, it will be apparent to those of skill in the art that any similar system or other effectively equivalent detection method can also be used.

Methods for evaluating the hybridization results vary with the nature of the specific probes used, as well as the controls. Simple quantification of the fluorescence intensity for each probe can be determined. This can be accomplished simply by measuring signal strength at each location (representing a different probe) on the high density array (*e.g.*, where the label is a fluorescent label, detection of the florescence intensity produced by a fixed excitation illumination at each location on the array).

One of skill in the art, however, will appreciate that hybridization signals will vary in strength with efficiency of hybridization, the amount of label on the sample nucleic acid and the amount of the particular nucleic acid in the sample. Typically nucleic acids present at very low levels (*e.g.*, < 1 pM) will show a very weak signal. At some low level of concentration, the signal becomes virtually indistinguishable from background. In evaluating the hybridization data, a threshold intensity value can be selected below which a signal is counted as being essentially indistinguishable from background.

The terms "background" or "background signal intensity" refer to hybridization signals resulting from non-specific binding, or other interactions, between the labeled target nucleic acids and components of the oligonucleotide array (*e.g.*, the oligonucleotide probes, control probes, the array substrate, *etc.*). Background signals may also be produced by intrinsic fluorescence of the array components themselves. A single background signal can be calculated for the entire array, or a different background signal may be calculated for each target nucleic acid. In a preferred embodiment, background is calculated as the average hybridization signal intensity for the lowest 5% to 10% of the probes in the array, or, where a different background signal is calculated for each target allele, for the lowest 5% to 10% of the probes for each allele. However, where the probes to a particular allele hybridize well and thus appear to be specifically binding to a target sequence, they should not be used in a background signal calculation. Alternatively, background may be calculated as the average hybridization signal intensity produced by hybridization to probes that are not complementary to any sequence found in the sample (*e.g.*, probes directed to nucleic acids of the opposite sense or to genes not found in the sample, such as bacterial genes where the sample is mammalian nucleic acids). Background can also be calculated as the average signal intensity produced by regions of the array that lack any probes at all. In a preferred embodiment, background signal is reduced by the use of a detergent (*e.g.*, C-TAB) or a blocking reagent (*e.g.*, sperm DNA, cot-1 DNA, etc.) during the hybridization to reduce non-specific binding. In a particularly preferred embodiment, the hybridization is performed in the presence of about 0.5 mg/ml DNA (*e.g.*, herring sperm DNA). The use of blocking agents in hybridization is well known to those of skill in the art (*see, e.g.*, Chapter 8 in P. Tijssen, *supra*).

The high density array can include mismatch controls. In a preferred embodiment, there is a mismatch control having a central mismatch for every probe in the array, except the normalization controls. It is expected that after washing in stringent conditions, where a perfect match would be expected to hybridize to the probe, but not to the mismatch, the signal from the mismatch controls should only reflect non-

specific binding or the presence in the sample of a nucleic acid that hybridizes with the mismatch. Where both the probe in question and its corresponding mismatch control show high signals, or the mismatch shows a higher signal than its corresponding test probe, there is a problem with the hybridization and the signal from those probes is ignored. For a given marker, the difference in hybridization signal intensity ($I_{allele1}$ - $I_{allele2}$) between an allele-specific probe (perfect match probe) for a first allele and the corresponding probe for a second allele (or other mismatch control probe) is a measure of the presence of or concentration of the first allele. Thus, in a preferred embodiment, the signal of the mismatch probe is subtracted from the signal for its corresponding test probe to provide a measure of the signal due to specific binding of the test probe.

The concentration of a particular sequence can then be determined by measuring the signal intensity of each of the probes that bind specifically to that gene and normalizing to the normalization controls. Where the signal from the probes is greater than the mismatch, the mismatch is subtracted. Where the mismatch intensity is equal to or greater than its corresponding test probe, the signal is ignored (*i.e.*, the signal cannot be evaluated).

For each marker analyzed, the genotype can be unambiguously determined by comparing the hybridization patterns obtained for each of the two labels, *e.g.*, color tags employed (Fig. 8). If hybridization is indicated for one color tag to its corresponding allele-specific probe (*e.g.*, "A") but not for the other color tag (*e.g.*, "G") (pattern at left in Fig. 8), then the indicated genotype of a diploid organism would be homozygous A/A. If hybridization is indicated only for the other color tag to its corresponding allele-specific probe (*e.g.*, "G") (pattern at center in Fig. 8), then the indicated genotype of a diploid organism would be homozygous G/G. If hybridzation is indicated for both color tags to their corresponding allele-specific probes (pattern at right in Fig. 8), then the indicated genotype of a diploid organism would be heterozygoous (A/G).

Marginal detection of hybridization, indicated by an intermediate positive result (*e.g.*, less than 1%, or from 1-5%, or from 1-10%, or from 2-10%, or from 5-10%, or from 1-20%, or from 2-20%, or from 5-20%, or from 10-20% of the average of all

positive hybridization results obtained for the entire array) may indicate either cross-hybridization or cross-amplification, depending on the overall hybridization pattern as indicated in Fig. 8. However, these can be distinguished by the unique pattern observed. Further procedures for data analysis are disclosed in U.S. Application 08/772.376, previously incorporated for all purposes.

HuSNP and other marker-specific arrays have been designed and used in genetic studies[9-10]. But the method developed in this study provides several advantages in dealing with many different genetic applications: (1) arrays based on a single generic design can be used to genotype different sets of genetic markers because no specific customized genotyping array is needed; (2) the pre-selected probe sequences synthesized on the tag array help ensure good hybridization results; (3) accurate quantitative measurement of the allele frequency in the tested samples can be achieved. Thus, reliable genotype results can be obtained not only for individual samples, but also for pooled samples. Besides SBE, other assays can be coupled with tag array assay, for example, oligonucleotide ligation assay (OLA)[19-21], invasive cleavage of oligonucleotide probes assay[22], allele specific PCR[23-24].

Our current tag chip contains over 32,000 unique tag probes. For most of the genetic application, for example, detecting mutations in one particular gene, it doesn't need such high-density chip. Therefore, smaller chips with fewer tags on the chip are sought after. Alternatively, multiple tags corresponding to one particular marker can be designed as to build the redundancy to the assay to assure accurate genotyping. Or multiple sets of tags for one set of SNPs can be designed, thus multiple samples can be processed and analyzed with one chip. Our current assay uses a two-color labeling scheme. But a four-color labeling/scanning system should warrant the assay can be done in a single tube reaction.

For broader genetic applications, for example, a study needs to genotype 100s to 1000s genetic markers, amplifying the genetic loci with multiplexing PCR is still the best strategy. However, to genotype 1000s to 10,000s markers, pre-amplification of the interested genetic loci will be very labor-intensive and costly. A whole-genome

approach should be explored, for example, strategies involved using total human genomic DNA directly, or genomic DNA amplified using some general amplification methods, *e.g*,. primer-extension preamplification, PEP[25], or total cDNA. In fact, we have tried to use total human genomic DNA directly as the SBE template in our tag array assay. 24 out of the 38 of the markers that we tested gave good signals (data not shown). Nevertheless, large amount of work are warranted as to solve both the sensitivity (signal intensity) and specificity (mis-priming) problems before the whole-genome approach become really useful.

The invention will be further illustrated by the following non-limiting examples. The content of references cited herein is incorporated herein by reference in its entirety.

EXEMPLIFICATION

METHODS

Collection and Isolation of DNA From Samples

DNA samples were collected by GenNet as part of the ongoing Family Blood Pressure Program. Samples were collected with consent and IRB approval in both Tecumseh, MI and Loyola, IL FAMILIES. Ascertainment was based on identification of a proband in the top 15th (Tecumseh) or 20th (Loyola) percentile of the community's blood pressure distribution. Full phenotypic information was obtained for each individual. DNA was extracted from 5-10 ml of whole blood taken from each individual using the standard "salting-out" method (Gentra Systems).

Primer Design

For each SNP, primary PCR amplification primers were designed as described previously[9]. The SBE primer was designed in a manner that its 3' terminates one base before the polymorphic site. Primer 3.0 software package (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi) was modified and used to pick SBE primers with batch sequences, at a predicted length of 20 (ranging from 18 to 26) nucleotide and melting temperature of 60°C (ranging from 54°C to 64°C). The SBE

primers were always picked from the forward direction first (i.e. 5' to the polymorphic site). If the SBE primer can't be picked from the forward direction, reverse direction is tried.

Multiplexing PCR

Specific genomic regions containing the 144 SNPs were amplified with 9 multiplex PCR reactions, each contains 50 ng of human genomic DNA, 0.1 μM of each primer, 1 mM deoxynucleotide triphosphates (dNTPs), 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 5 mM $MgCl_2$ and 2 units of AmpliTaq Gold (Perkin Elmer) in a total value of 25 μl. PCR was performed on a Thermo Cycler (MJ Research), with initial denaturation of the DNA templates and Taq enzyme activation at 96°C for 10 minutes; followed by 40 cycles of denaturation at 94°C for 30 seconds, 57°C for 40 seconds, and 72°C for 1 minute and 30 seconds; and the final extension at 72°C for 10 minutes.

SBE Template Preparation

1 μl of Exonuclease I (Amersham Life Science, 10 U/μl) and 1 μl of Shrimp Alkaline Phosphatase (Amersham Life Science, 1 U/μl) were added to a 25 μl PCR products (see above), and incubated at 37°C for 1 hour. The enzyme activities were inactivated at 100°C for 15 minutes. The enzymatically treated samples were applied to a S-300 column (Pharmacia), as to further reduce the residual PCR primers and dNTPs, and replace the buffer with ddH_2O.

Multiplexing SBE Reaction

SBE is carried out in a 33 μl reaction, using 6 μl of the template (see above), 1.5 nM of each SBE primer, 2.5 units of Thermo sequenase (Amersham), 52 mM Tris-HCl (pH 9.5), 6.5 mM $MgCl_2$, 25 μM of fluorescein-N6-ddNTPs (NEN), 7.5 μM biotin-N6-ddUTP or biotion-N6-dCTP, or 3.75 μM biotin-N6-ddATP, and 10 μM the other cold ddNTPs.

Extension reaction was carried out on a Thermo Cycler (MJ Research), with 1 cycle of 96°C for 3 minutes, then 45 cycles of 94°C for 20 seconds and 58°C for 11 seconds.

After SBE reaction, 9 reactions from each sample were combined and mixed with 30 µl of 100 µg/ml glycogen (Boehringer Mannheim), 18.75 µl of 8 M LiCl (Sigma), and 1125 µl of pre-chilled (-20°C) ethanol (Abs.), and precipitated by centrifugation at the top speed (Eppendorf centrifuge 5415C) for 15 minutes at room temperature; precipitated samples were dried at 40°C for 40 minutes and re-suspended in 33 µl ddH_2O.

Tag Array Design and Hybridization

For each tag sequence, two probes were synthesized on the array. One is exactly the designed tag sequence (referred to as a Perfect Match, or PM probe). The other one is identical except for a single base difference in a central position (referred to as a Mismatch, or MM probe). The mismatch probe services as an internal control for hybridization specificity. Over 32,000 20-mer tag probes (and their companions) were chosen[11] and fabricated on a 8 mm x 8mm size of array. Each probe (feature) occupies a 30 microns x 30 microns area. The sets of arrays were synthesized together on a single glass wafer on which 100 arrays were made.

The labeled sample was denatured at 95°C - 100°C for 10 minutes and snap cooled on ice for 2 - 5 minutes. The tag array was pre-hybridized with 6 X SSPE-T (0.9 M NaCl, 60 mM NaH_2PO_4, 6 mM EDTA (pH 7.4), 0.005% Triton X-100) + 0.5 mg/ml of BSA for a few minutes, then hybridized with 120 µl hybridization solution (as shown below) at 42°C for 2 hours on a rotisserie, at≈ 40 RPM. Hybridization Solution consists of 3M TMACL (Tetramethylammonium Chloride), 50 mM MES ((2-[N-Morpholino]ethanesulfonic acid) Sodium Salt) (pH 6.7), 0.01% of Triton X-100. 0.1 mg/ml of Herring Sperm DNA, 50 pM of fluorescein-labeled control oligo, 0.5 mg/ml of BSA (Sigma) and 29.4 µl labeled SBE products (see below) in a total of 120 µl reaction.

The chips were rinsed twice with 1X SSPE-T for about 10 seconds at room temperature, then washed with 1X SSPE-T for 15 - 20 minutes at 40°C on a rotisserie, at ≅ 40 RPM. And then wash the chip 10 times with 6X SSPE-T at 22°C on a fluidic station (FS400, Affymetrix). The chips were stained at room temperature with 120 µl staining solution (2.2 µg/ml streptavidin R-phycoerythrin (Molecular Probes), and 0.5 mg/ml acetylated BSA, in 6 x SSPET) on a rotisserie for 15 minutes, at ≅ 40 RPM. After staining, the probe array was washed 10 times again with 6 x SSPET on the FS400 at 22°C. The chips were scanned on a confocal scanner (Affymetrix) with a resolution of 60-70 pixels per feature, and two filters (530-nm and 560-nm, respectively). GeneChip Software (Affymetrix) is used to convert the image files into digitized files for further data analysis.

Clustering Analysis

For a given marker (at a given tag probe position), the intensity of each of the two colors (fluorescein and phycoerythrin) was calculated as the intensity at the perfect match position (PM) minus that at the mis-match position (MM). Negative fluorescein or phycoerythrin intensity values are treated as if they were zero. The Phat values were computed as the ratio of the intensities (fluorescein/fluorescein + phycoerythrin). The Phat values were sorted, and the optimal set of ranges for AA, AB and BB genotypes given the hypothesis of 2 or 3 clusters was considered, subject to the following rules: at most 4 points (outliers) may be excluded from the genotype ranges. For 2 groups, the total range Phat values must be at least 0.3. For 3 groups, the total range Phat values must be at least 0.5. Ranges must be separated by a gap of at least 0.1. The width of a range may be at most 0.4. A score was then computed as: Score = 1 - (sum of range widths / total range) – (outliers * 0.1).

The set of ranges with the best score was found and used to call genotypes. This score increases with narrow ranges, while decreases with the number of points that are left out of any range. Therefore, it tends to be optimal when all the phat values are contained within relatively small ranges.

ABI Sequencing to Determine Genotypes

To independently confirm the genotypes called from the tag array assay, three samples (904957000000, 904896000000, and 904889000000) were sequenced using gel-electrophoresis based method. Samples were amplified for all sites with T7 and T3 tagged primers, using standard PCR cycling conditions (2.5 µl of 20 ng/µl DNA, 0.375 µl of 20 µM primer (X2), 1.5 µl of 10X PCR buffer, 0.9 µl 25mM Mg^{2+}, 0.15 µl 10mM dNTPs, 0.25 µl 10 U/µl Taq DNA Polymerase (Sigma), brought up to 15 µl with ddH_20 per tube). Some products were sequenced directly, while a M13 nesting strategy was used due to the close proximity of the polymorphic base to the primer end. Samples from the initial amplification were diluted 1:50 with ddH_20, and amplified with M13F-T7 (TGTAAAACGACGGCCAGTTAATACGACTCACTATAGGGAGA; SEQ ID NO: 9) and M13R-T3 (AACAGCTATGACCATGAATTAACCCTCACTAAAGGGAGA; SEQ ID NO: 10) primers using standard PCR conditions. All PCR products were cleaned with Exonuclease I (Amersham 0.15 µl of 10 U/µl per well) and Shrimp Alkaline Phosphatase (Amersham, 0.30 µl of 1 U/µl per well) in a volume of 10 µl. Dye terminator sequencing using a M13R primer (AACAGCTATGACCATG; SEQ ID NO: 11) or T7 primer (TAATACGACTCACTATAGGGAGA; SEQ ID NO: 12) on an ABI377 (Perkin Elmer) using Big Dyes (Perkin Elmer) was performed to determine the genotype status for each SNP in all three individuals. Trace files were read with Edit View 1.0 (Perkin Elmer) software.

EXAMPLE 1

DNA from a individual is isolated, and amplified with primers from 15 previously-characterized (i.e., known) SNPs. Amplification is allowed to proceed as described in Hudson, T.J. *et al.* (Science 270:1945-1954 (1995)) and Dietrich *et al.* (Dietrich, W. F. *et al.*, Nature 380:149-152 (1996); Dietrich, W. F. *et al.*, Nature Genetics 7:220-245; Dietrich, W. *et al.*, Genetics 131:423-447 (1992)). For example, in a 50 µl reaction volume, 0.5 ng of template nucleic acid/target polynucleotide is added

to 1 μM forward amplification primer, 1 μM reverse amplification primer, 200 μM dGTP, 200 μM dTTP, 200 μM dATP, 3.5 mM $MgCl_2$, 1.0 mM Tris-HCl (pH 8.3), 50 mM KCl, 0.02 μM molecular probe, and 0.25 units of polymerase enzyme. The reaction mixture can then be subjected to a two-step amplification process, performed on a Tetrad (MJ Research, Watertown, Massachusetts), with the conditions: denaturation at 94°C for 60 seconds, followed by an annealing/extension step at 53°-56°C for one minute. The denaturation and annealing/extension steps are repeated for 40 cycles. Alternatively, a three-step thermocycling reaction can be used, such as 94°C for 60 seconds, followed by annealing at 53°-56°C for 30 seconds, followed by extension at 72°C for one minute the three steps being repeated for 40 cycles. This may be followed by an optional extension step at 72°C for five minutes.

After amplification is complete, locus-specific tagged oligonucleotides specific for the 10 SNPs are added, and are allowed to hybridize to the amplification products.

Reagents for a single base extension reaction are then added, where each of the four ddNTPs is labeled with a different fluorophore. Single base extension is then performed as described by Kobayashi et al. (Mol. Cell. Probes 9:175-182 (1995)).

After the reaction is complete, the reaction products are placed in contact with the universal array, and the reaction products allowed to hybridize, each product to its appropriate oligonucleotide tag on the array. The chip is then assayed in a fluorometer, and the wavelength emitted at each address in the array is recorded. From this data, the genotype at each individual SNP is determined.

EXAMPLE 2

Two alleles of template were mixed at ratios of 1:30, 1:10, 1:3, 1:1, 3:1, 10:1, and 30:1. These were labeled with different color labels by single-base extension reaction and hybridized to a tag array. A correlation was observed between the signal intensity ratio and the template concentration ratio over a 900-fold dynamic range. See Figure 2.

EXAMPLE 3

A set of tag sequences is selected such that the tags are likely to have similar hybridization characteristics and minimal cross-hybridization to other tag sequences. An oligonucleotide array of all of the tags is fabricated. The design and use of such a 4,000-20mer-tag array for the functional analysis of the yeast genome has been described (1). More recently, Affymetrix designed and fabricated an array with a set of more than 16,000 such tags. The tag sequence synthesized on the chip can be 20-mer, 25-mer, or other lengths.

EXAMPLE 4

Marker specific primers are used to amplify each genetic marker (*e.g.* SNP). A multiplex PCR strategy is used to amplify these markers from genomic DNAs of tested individuals (2). After PCR amplification, excess primers and dNTPs are removed enzymatically. These enzymatically treated PCR products then serve as templates in the next SBE reaction. Please note that these templates (PCR products) are double stranded, which are different from the templates used in other protocols (3, 4). For example, in Minisequencing (3) and Genetic Bit Analysis (GBA, 4), a double stranded template has to be converted to a single stranded template prior to the base extension reaction. The methods used for this conversion are costly, laborious, and hard to automate.

EXAMPLE 5

In the protocol described below, an SBE primer is designed for each genetic marker which terminates 1 base before the polymorphic site. However, other primer design schemes can be used. The primer for each marker is tailed with an unique tag which is complementary to a specific probe sequence synthesized on the tag chip. The extension reaction is multiplex, in which SBE primers corresponding to multiple markers were added in a single reaction tube, and extended in the presence of pairs of

ddNTPs labeled with different fluorophores, *e.g.* for an A/C variant. there might be a ddATP-red and DDCTP-green.

EXAMPLE 6

The resulting mixture is hybridized to the tag array. Each tag corresponds to a single marker. The ratio of the intensities of the colors indicates the genotype (or the allele frequency, ranging from 0% to 100%) of the samples tested.

EXAMPLE 7

SBE template preparation: Marker specific primers are used to amplify each single nucleotide polymorphism (SNP). A multiplex PCR strategy is used to amplify these SNPs (Science 280:1077-1082, 1998).

Multiplex PCR:

Multiplex PCR reaction is carried out with AmpliTaq Gold and 25 primer pairs in a 25µl reaction volume. SNPs with same base composition at the polymorphic site (i.e. A/G, T/C, etc) are pooled together.

PCR reagents:

10XPCR Multiplex Buffer (II): 100 mM Tris/HCl (pH 8.3)
500 mM KCl

25 mM dNTPs

F & R Primers (for each primer, the conc. is 1 µM)

20 ng/µl Genomic DNA

Multiplex PCR reaction (25 ul)

Primer Mix (1 µM each) 2.5 µl

Genomic DNA (20 ng/µl) 2.5 µl

10XPCR Buffer II	2.5 µl
25 mM $MgCl_2$	5 µl
25 mM dNTPs	1 µl
AmpliTaq Gold (5U/µl)	0.4 µl
ddH_2O	up to 25 µl

PCR conditions

96°C	10 min

40 cycles :

94°C	30 sec
57°C	40 sec
72°C	1 min 30 sec

72°C	10 min
4°C	O/N

Enzymatic treatment of PCR products to degrade and de-phosphorylate the unused primers and dNTPs, respectively:

To a 25 µl PCR products, add 1 µl of Exonuclease I (Amersham Life Science, 10 U/µl) and 1 µl of Shrimp Alkaline Phosphatase (Amersham Life Science, 1 U/µl), and incubate at 37° C for 1 hour. Inactivate the enzyme activities at 100°C for 15 minutes. Apply the sample to a S-300 column (Pharmacia), to further reduce the residual PCR primers and dNTPs, and replace the buffer with ddH_2O. The sample is ready for next SBE reaction.

Single Base Extension (SBE):

An SBE primer is designed for each SNP which terminates 1 base before the polymorphic site. The primer for each SNP is tailed with a unique tag which is complementary to a specific probe sequence on the tag chip. The SBE reaction is also multiplexed at 25-plex.

Reaction Mixture (33 µl):

Template (see above)	6 µl
SBE Primer mix (20 nM for each primer)	2.5 µl
5X Thermo Sequenase buffer	6.6 µl
Bio-(d)dNTP(X nmol/µl*, NEN)	0.5 µl
Flu-ddNTP(1nmol/µl, NEN)	0.8 µl
Other two cold –ddNTPs(1nmol/µl, Biopharmacia)	0.3 µl each
Thermo Sequenase(6.4 U/µl) (Amersham)	0.4 µl
ddH_2O	up to 33µl

* X= 0.5 when it is Bio-ddUTP or bio-dCTP(0.5 mM), or X= 0.25 when it is bio-ddATP (0.25 mM)

PCR program:

96°C	3'	1 cycle
94°C	25"	
58°C	11"	45 cycles
4°C forever		

Precipitation:

After SBE reaction, we combined 9 tubes for each sample, mix with 30 µl of 100 µg/ml glycogen (Boehringer Mannheim), then precipitated with 18.75 µ l of 8 M LiCl, and 1125 µl of pre-chilled (-20°C) ethanol (Abs.). Mix well; then centrifuge at the top speed (Eppendorf centrifuge 5415C) for 15 min at room temperature; Decant the supernatant, and dry the samples at 40C for 40 min, re-suspend the samples in 33 µl ddH2O, now it is ready for hybridization.

Hybridization:

The prepared sample is denatured at 100°C for 10 minutes and snap cooled on ice for 2-5 minutes. The universal tag chip is pre-hybridized with 6 X SSPE-T (0.9 M NaCl, 60 mM NaH_2PO_4, 6 mM EDTA (pH 7.4), 0.005% Triton X-100) + 0.5mg/ml of BSA, then hybridized with 120 µl hybridization solution (as shown below) at 42°C 2 hours on a rotisserie, at≈ 40 RPM.

The hybridization solution contains:

5M TMACL	72 µl
0.5M MES (pH 6.7)	12 µl
1% Triton X-100	1.2 µl
HS DNA (10mg/ml)	1.2 µl
Flu-c213 (5 nM)	1.2 µl
BSA (20 mg/ml)	3.0 µl

Plus 29.4 µl prepared sample (see above).

Post-Hybridization Wash:

Rinse the chip with 1X SSPE-T 10" twice first, then wash with 1X SSPE-T for 15-20min at 40°C on a rotisserie, at ≈ 40 RPM. And then wash on a fluidic station (FS400, Affymetrix) 10 times with 6 x SSPET at 22°C.

Staining:

Stain the chip at room temperature with 120 µl staining solution (2.2 µg/ml streptavidin R-phycoerythrin (Molecular Probes), and 0.5 mg/ml acetylated BSA, in 6 x SSPET) on a rotisserie for 15 minutes, at≈ 40 RPM. After staining, the probe array was washed 10 times again with 6 x SSPE-T on the FS400 at 22°C.

Scanning:

The chips were scanned on a confocal scanner (Affymetrix) with a resolution of 60-70 pixels per feature, and two filters (530-nm and 560-nm, respectively). GeneChip Software (Affymetrix) is used to convert the image files into digitized files for further data analysis.

EXAMPLE 7

Genotyping With High-Density Oligonucleotide "Tag" Arrays

A genotyping method based on the use of a high-density "tag" array that contains over 32,000 pre-selected 20-mer oligonucleotide probes, combined with marker-specific PCR amplifications and single base extension (SBE)[1-2] reactions has been developed. We have used this method to genotype a collection of 144 single-nucleotide polymorphism (SNPs) identified from 49 hypertension candidate genes[3]. First, marker-specific primers were used in multiplex PCR reactions to amplify specific genomic regions containing the SNPs. The PCR amplified DNA products were then used as templates in SBE reactions. Each SBE primer comprises a 3' portion and a 5' portion. The 3' portion is complementary to the specific SNP locus and terminates one base before the polymorphic site. The 5' portion comprises a unique sequence, which is complementary to a specific oligonucleotide probe synthesized on the "tag" array. The extension reaction is multiplex, with SBE primers corresponding to multiple SNPs in a single reaction tube. The primers are extended in the presence of two-color labeled ddNTPs, and the resulting mixture is hybridized to the tag array. The intensity ratio of the two colors was used to deduce the genotypes of the samples tested.

The tag array strategy begins with an array of tag sequences selected in a manner that all tag probes are in the same length, *e.g.* 20-nucleotide long, with similar melting temperature and G-C content, and the lowest sequence homologous among each other[11]. Therefore, these tags are likely to have similar hybridization characteristics and minimal cross-hybridization to other tag sequences.

The design and use of a 4,000-tag array for the functional analysis of yeast *Saccharomyces cerevisiae* genes[11] and drug sensitivity studies[12] have been described. More recently, we have designed and fabricated an array that contains more than 32,000 such tags, and developed it as a genotyping tool, in combination with marker-specific PCR amplifications and SBE reactions.

As shown in Fig. 7, marker specific primers are designed and used to amplify each single nucleotide polymorphism (SNP). A multiplex PCR strategy is used to amplify these SNPs from genomic DNAs[9]. In general, SNPs with same base composition at the polymorphic site (*e.g.* all the A/G polymorphisms) are grouped together. After PCR amplification, excess primers and dNTPs are degraded and de-phosphorylated using Exonuclease I and Shrimp Alkaline Phosphatase, respectively. These enzymatically treated PCR products (double-stranded) are then served as templates in the SBE reaction. A SBE primer is designed for each genetic marker, which terminates one base before the polymorphic site. Each primer is tailed with a unique tag that is complementary to a specific probe sequence synthesized on the tag array. The extension reaction is multiplex, in which SBE primers corresponding to multiple markers (up to 56 markers that we have tested so far) were added in a single reaction tube, and extended in the presence of pairs of ddNTPs labeled with different fluorophores, *e.g.* for an A/G variant, biotin-labeled ddATP and fluorescein-labeled ddGTP are used. The resulting mixture of SBE reactions is hybridized to the tag array. Each tag hybridizes to a specific probe position on the chip. The ratio of the intensities of the colors indicates the genotype (homozygous wild type, or homozygous mutant, or heterozygous) or the allele frequency (ranging from 0% to 100%) in the samples tested.

In a comparison of the results of using single-stranded and double-stranded PCR products as the templates in the current SBE/tag array assay, no significant difference was found (data not shown). However, in previously published protocols of minisequencing[13-15] and genetic bit analysis[16-18], a double-stranded template has to be converted to a single-stranded template prior to the base extension reaction. The methods used for this conversion were costly, laborious, and hard to automate.

The tag array assay provides a fairly accurate quantitative measurement of the allele frequency in samples tested. As shown in Figure 2, we have synthesized two artificial SBE templates. They are identical, except the 21st position: T in template-T, and G in template-G. We then mixed the two templates at ratios of 1:10, 1:3, 1:1, 3:1, 10:1, and 30:1, which is a 300-fold dynamic range. Six SBE primers, which have the same 3' portion (the portion complementing to the template sequence) but different 5' portion (the portion complementing to the tag probes on the tag arrays) were designed (Figure 2), and extended in the presence of the SBE templates mixed at different ratios, and biotin-labeled ddATP and fluorescein-labeled ddCTP. As shown in Fig. 8, the intensity ratio of the two colors and the template concentration ratio (i.e. the allele frequency) appears to form a fairly good linear correlation in the 300-fold dynamic range that we tested.

To further test the robustness and the efficiency of the tag array/SBE assay method for genotyping application, we set out to type a portion of the SNPs that we had identified from a large-scale polymorphism screening study with the hypertension candidate genes[3]. Initially, we selected 173 SNPs from 56 hypertension candidate genes. These SNPs were chosen for their being occurred in promoter regions, or splicing junctions, or coding regions in which the nucleotide changes caused amino acid changes. We reason that these SNPs can be the good candidates for being the functional mutations predisposed to the disease. Therefore, the assay developed in this study could then be used in large-scale association studies in hypertension. PCR primers were designed and tested individually for these 173 SNPs. 8 of them (4.6%) failed to amplify. SBE primers were then designed for the remaining 165 SNPs. A multiplexing PCR and

multiplexing SBE assay was developed with a complexity of 9 to 28 markers in each reaction and a total of 9 reactions for the 165 markers. 21 of them (12.7%) failed in the multiplexing PCR and multiplexing SBE assay. Therefore, 144 markers from 49 genes passed the assay development. The gene location, polymorphic sites, and the designed primers for these 144 markers were summarized in Table 1.

We then genotyped 44 individuals using 44 tag arrays. Good hybridization signals were obtained in 96.5% (6116 / 6336 (144 x 44)) of the cases. The signal intensity values from the hybridization results were used in clustering analysis for each of the 144 markers. Genotypes for each individual at the 144 loci were assigned automatically based on the clustering results, with some manual editing. Data Desk 6.0 (Data description, Inc.) was used to manually display the clustering analysis results (of the intensity ratios of the two colors). Overall, 80-85% of the markers form good cluster(s).

We have performed the gel-based DNA sequencing to determine the genotypes at 115 loci in 3 of the 44 individuals (see Methods). Comparison of the ABI sequencing results and the chip results resulted in 14 discrepancies (4%), out of 115 x 3 = 345 genotype calls. Most of the discrepancies occurred in cases where one method called homozygous, while the other method called heterozygous. In one case (marker ICAM1ex6.254), where the ABI sequencing method called G/G, but the tag array /SBE assay method called A/A in all the three individuals, we believe the discrepancies are due to mis-priming of the SBE primer to adjacent sequences.

We also tested the reproducibility of the tag array/SBE assay genotyping method. We repeated the multiplexing PCR, SBE and the chip hybridization experiments in 4 individuals. The ratios of the two colors (for each of the 144 markers) in the replicated experiments are not all exactly the same, but they all fall into the same cluster (i.e. giving the same genotype call). Therefore, we didn't find any discrepancy in the genotyping call of duplicated samples.

TABLE 1

Gene/Exon/Position	SNP Flanking Sequence	Forward Primer	Reverse Primer	SBE Primer
AADDEX10.246	TTCCGAGGAA(G/T)GGCAGAATGG	GACGAAGCTTCCGAGGA	GGGACTGCTTCCATTCTGC	AGAGTCTATAAGCAT CGTCGGGCGACGAAG CTTCCGAGGAA
AADDEX13.173	CAGAAGGGCT(C/G)TGAAGGTGAG	GAGAGGAAGCAGAAGGGC	GACCACAAGCACTCACCTTC	TCAGACAATTCTATA CGCGGTGGAGAGGAA GCAGAAGGGCT
ACEEX13.138	TGCTGGTCCC(C/T)AGCCAGGAGG	GCACCCTCTGCTGGTCC	TGACTGTCACCTGTTGGGA	TCGTGAGTTGTCCTGC TGCAGCACCCTCTGC TGGTCCC
ACEEX13.151	CCAGGAGGCA(T/C)CCCAACAGGT	GCACCCTCTGCTGGTCC	TGACTGTCACCTGTTGGGA	GCCTGTAATGGTGGA TCTCAGTCCCCAGCC AGGAGGCA
ACEEX13.202	AACAACCAGC(A/G)GCCAGACAAC	AGCCAGGCAACAACCAG	GTGGGTGGTTGTCTGGC	GATCTGTCTGACGCT GTATGGCAGCCAGGC AACAACCAGC
ACEEX15.144	AACGGGCAGC(G/A)CTGCCTGCCC	AGGACCTAGAACGGGCAG	TCCTGGGCAGGCAGC	CGTGATAATGCGTCT CGTAGCAGGACCTAG AACGGGCAGC
ACEEX17.19	AGCCATTCAA(C/A)CCCCTACCAG	TGGAGCTCAAGCCATTCA	CGTCAGATCTGGTAGGGGG	CATTATCGGACATGC TCACTTGGAGCTCAA GCCATTCAA
ACEEX17.52	TGATGGCCAC(A/G)TCCCGGAAAT	GACGAATGTGATGGCCA	GGTCTTCATATTTCCGGGAT	ATGATGAGCCGTGAT GACCCCTGACGAATG TGATGGCCAC
ACEEX18.130	CACTCTACCT(C/G)AACCTGCATG	GAGCTGCAGCCACTCTACC	CGTAGGCATGCAGGTTG	TACATCGCTTGCATG AGTGTGAGCTGCAGC CACTCTACCT

ACEEX21.150	CATGAGGCCA(T/C)TGGGGACGTG	CGGCTTCCATGAGGCC	GGCTAGCACGTCCCCAA	GATCTGGCTTCAACT GTATGCCGGCTTCCA TGAGGCCA
ACEEX22.19	TGACATCAAC(T/G)TTCTGATGAA	TTGCAGAGCATGACATCAA	AAGGGCCATCTTCATCAGA	TGCCTAGCTTTCCATA TCGGCCTTGCAGAGC ATGACATCAAC
ACEEX24.118	CCAAGGAGGC(C/T)GGGCAGCGCC	CATCTACCAGTCCAAGGAG G	TCACCCAGGCGCTGC	TATCTCGCTTGCTATC AACGATCTACCAGTC CAAGGAGGC
ACEEX24.16	CCAGGTACTT(T/C)GTCAGCTTCA	TCGCTCTGCTCCAGGTACT GA	GGAACTGGATGATGAAGCT	GCCTAAGCTCTGTCG CTGATTCGCTCTGCTC CAGGTACTT
ACEEX26.154	CTCAGCCAGC(G/A)GCTCTTCAGC	CTGGGCCTCAGCCAG	GCGGATGCTGAAGAGCC	TCTATTGCTGTTCGGC GGCAACCCTGGGCCT CAGCCAGC
ACEEX26.174	CATCCGCCAC(C/A)GCAGCCTCCA	TCTTCAGCATCCGCCA	GCCGGTGGAGGCTGC	AGCAGAGATGGACAG ACCTCCTCTTCAGCAT CCGCCAC
ACEEX26.205	CACGGGCCCC(A/C)GTTCGGCTCC	CACTCCCACGGGCCC	CACCTCGGAGCCGAACT	GCTGGCGGTTCATGC AATCTTCCACCTCGG AGCCGAAC
ACEEX26.224	CCGAGGTGGA(G/A)CTGAGACACT	TCGGCTCCGAGGTGG	CACCTCAGGAGTGTCTCAGC	TATCTGCGTTGCTGAC GTGCCAGTTCGGCTC CGAGGTGGA
ACEEX8.106	AGGATCTGCC(C/T)GTCTCCCTGC	CCTGCAGTACAAGGATCTG C	CCCGACGCAGGGAGA	GATCCGTATGTCGAA TGGCTCTGCAGTACA AGGATCTGCC
ACEP.-3892	TAAGGGGGGG(T/C)TGCTGTACAT	CCACTGAGGATAAGGGGG	GAAGATATTTGCAAAGTAT GTACAGC	CCAGAGGTGCGGTCA CATATCACTGAGGAT AAGGGGGGG

ACEP.-5466	TATAGTATAT(A/C)TATGCCCAGC	GTCATGCCATGTCACATATA TTATAGT	GACCATGGCTGGGCAT	GCATCTTCGCCAGCT ATATTGGTTGACCAT GGCTGGGCATA
ACEP.-93	GCTGGGGACT(T/C)TGGAGCGGAG	AGGAACCTCGGCCCG	GCTTCCTCCTCCGCTCC	CACTTACGGCCATGC TGAATCCCGCGCCGC TGGGGACT
ADDBEX15.85	AGTTCTTCAG(C/T)GTTGCCCTCC	CCGTGTGCGAGTTCTTCA	CCAGATGTGGAGGGCAA	CACTGTACGCACTGG AGCTACGTGTGCGAG TTCTTCAG
ADDBEX3.138	CTCAGAGGAC(G/A)ACCCCGAGTA	TGACCGCTTCTCAGAGGA	GCGCATGTACTCGGGG	GTGTGCATTGAGTCT ATGACTTTGACCGCTT CTCAGAGGAC
ADRB3EX1.416	GGCCATCGCC(T/C)GGACTCCGAG	CATCGTGGCCATCGC	CTGGAGTCTCGGAGTCCA	CGTCTCATGCCTGCGT ATAGTGGTCATCGTG GCCATCGCC
ADROMEX3.81	GGATGTCCAG(C/G)AGCTACCCCA	GGAACTGCGGATGTCCA	GCCCGGTGGGGTAGC	TACATCATTGCGAGT CATGGAAGAGGGAAC TGCGGATGTCCAG
AEIEX14.159	CTTCTTTGCC(A/T)GATGCTGCG	CCGGTACCTTCTTCTTTGC	TGAACTTGCGCAGCATC	ATACGCTCTGCCATA CGTGAGCCGGTACCT TCTTCTTTGCC
AEIEX4.36	TCAGCTCACG(A/C)CACCGAGGCA	CCGACCTCTGGTTTTCAGC	TGGCTGTTGCCTCGGT	TTGCGCCATTTGGAC ATGCTACCTCTGGTTT TCAGCTCACG
AEIEX4.89	GGGTACCCAC(A/G)AGGTGAGGAC	CACACCCGGGTACCCA	GGCTGGGGTCCTCACC	GCCTGATATTCATTCA CAGCACATCACACCC GGGTACCCAC
AGT.385	CCGTTTCTCC(T/C)TGGTCTAAGT	GACTTTGAGCTGGAAAGCA G	CATGCAGCACACTTAGACC A	CTGTCGTCTAGTCTCT GAGGCATGCAGCACA CTTAGACCA

AGTEX2.354	GGATGCTGGC(C/T)AACTTCTTGG	TGGTCGGGATGCTGG	CGGAAGCCCAAGAAGTTG	TTTCGTGCTTTGGAG ACAGCAATGGTCGGG ATGCTGGC
AGTEX2.755	TTCACAGAAC(T/G)GGATGTTGCT	CGCTCTCTGGACTTCACAGA	TCTCAGCAGCAACATCCA	TGCCGTGTTGGTGCTT CACACTCTCTGGACTT CACAGAAC
AGTEX2.827	TGCTCCCTGA(T/C)GGGAGCCAGT	AGACTGGCTGCTCCCTG	TCCACACTGGCTCCCA	TCGTCCACTTTAGCAT GATGAAGACTGGCTG CTCCCTGA
AGTEX5.376	GGAAAGCAGC(C/G)GTTTCTCCTT	GACTTTGAGCTGGAAAGCA G	CATGCAGCACACTTAGACC A	TACATACTTGCAGTG CGTTCACTTTGAGCTG GAAAGCAGC
AGTEX5.385	CCGTTTCTCC(T/C)TGGTCTAAGT	GACTTTGAGCTGGAAAGCA G	CATGCAGCACACTTAGACC A	CGTCGTGCTGCGTGA CTATAGGAAAGCAGC CGTTTCTCC
AGTEX5.641	TCGGTTTGTA(T/G)TTAGTGTCTT	GCATTGCCTTCGGTTTGT	TCATGTTCTTACATTCAAGA CACTAAA	TGAGAGTCTGTTCTT AGGCCCATTTTTGCAT TGCCTTCGGTTTGTA
AGTEXP1.101	CTGTGCTATT(G/C)TTGGTGTTTA	CTTTCAATCTGGCTGTGCTA T	GGGGAGACTGTTAAACACC AA	TACATAATTGCCATG ACGGGTTCAATCTGG CTGTGCTATT
AGTEXP2.160	CCTTGGCCCC(G/A)ACTCCTGCAA	TGGGAACCTTGGCCC	ACCGAAGTTTGCAGGAGTC	GAGAATGCTGTGTATAG TGTCCTTTCTGGGAAC CTTGGCCCC
AGTEXP2.203	ACCCTGCACC(G/A)GCTCACTCTG	TGTGTAACTCGACCCTGCAC	CTGCTGAACAGAGTGAGCC	CGTCTCGCTGGTCACT AATGGTGTAACTCGA CCCTGCACC
AGTEXP2.35	CTGCACCTCC(G/A)GCCTGCATGT	TCTGCCCTCTGCACCTC	CAGGGACATGCAGGCC	GATCTCTGTGAAGTT AGTGCCCCTCTGCCCTC TGCACCTCC

AGTEXP3.144	TAAATAGGGC(C/A)TCGTGACCCG	CACCCCTCAGCTATAAATAG GG	CGGCAGCTTCTTCCCC	TATAAAGATTGCGGT CAGGCCCCTCAGCTA TAAATAGGGC
AGTEXP3.158	TGACCCGGCC(A/G)GGGGAAGAAG	CACCCCTCAGCTATAAATAG GG	CGGCAGCTTCTTCCCC	CCAGTCGGTGTAGCA GCAATTAGGGCCTCG TGACCCGGCC
AGTEXP3.173	AAGAAGCTGC(C/T)GTTGTTCTGG	GCCAGGGGAAGAAGCTG	GCTGTAGTACCCAGAACAA CG	GTGTGCTCTTCTCTCGCT GCAAGCCAGGGGAAG AAGCTGC
ALDREDEX1.162	CAAGATGCCC(A/T)TCCTGGGGTT	ACGGCGCCAAGATGC	AGGTACCCAACCCCAGG	ATACCGGCTGCTACA CAGTGAACGGCGCCA AGATGCCC
ALDREDEX1.71	GTACGCGCCG(C/G)GGCCAAGGCC	GCTATTTAAAGGTACGCGCC	TACGGTGCGGCCTTG	CAAATAGTGTGCGAG GATCTGCTATTTAAA GGTACGCGCCG
ALDREDEX3.28	TCGCTGGCTT(A/T)GCTGTGGTGC	GCCCTCTCGCTGGCT	CATGGTACGTGCACCACAG	TGAGACATTGTGCAA ATCGGACATGTGCCC TCTCGCTGGCTT
ANPEX3.33	TTTGCAGTAC(T/C)GAAGATAACA	ATATGTCTGTGTTCTCTTTG CAGT	CTCCCTGGCTGTTATCTTCA	GATAGCAGTTCACTA CCTGGGTCTGTGTTCT CTTTGCAGTAC
APOA2.249	TCCTGTTGCA(T/C)TCAAGTCCAA	TTGGAATCCTGCTTCCTGT	GATCTGAGGTCCTTGGACTT G	GGCATCACTGGTTAC GTCTGATCTGAGGTC CTTGGACTTGA
APOA4.3058	AGGAACAGCA(T/G)CAGGAGCAGC	AGCAACAGCAGGAACAGC	CACCTGCTCCTGCTGC	GTCTGACTTGAGTTA CATGGGAGCAACAGC AGGAACAGCA
APOCIEX1.462	TTCTGTCGAT(C/G)GTCTTGGAAG	TGGTGGTGGTTCTGTCGA	TCCCACTTTTACCTTCCAAG A	GGTCTTCCTATATGTG CGCGTCCTGGTGGTG GTTCTGTCGAT

APOC2.804	CTTTCTCCCC(A/T)GGGACTTGTA	ACCATCTGTGCTTTCTCCC	TCATGGCTGCTGTGCTT	TGAGAAGTTGTGAAG ATCCCTAACCATCTGT GCTTTCTCCCC
APOC2.819	CTTGTACAGC(A/C)AAAGCACAGC	ACCATCTGTGCTTTCTCCC	TCATGGCTGCTGTGCTT	GCCAGGCGTTCAGAT GCAATCCCAGGGACT TGTACAGC
APOC4.3162	CTGGGTCCGC(T/G)CACCAAGGCC	AGGGACCTGGGTCCG	AGGAACCAGGCCTTGGT	GCTGGTCGTGGTCCA ATCATTGAGGGACCT GGGTCCGC
APOER2EX12.68	ACTGTCCAGC(A/C)TTGACTTCAG	CAAGCTACACCAACTGTCC AG	TCTGTTGCCTCCACTGAAG	GACCATGCTGGCTTA CCTGTAAGCTACACC AACTGTCCAGC

AT2EX3.807	GGGAAGAACA(G/A)GATAACCCGT	GACGAATAGCTATGGGAAG AAC	ACTTGGTCACGGGTTATCC	TGGCATCGTTTCACCT GCTGGACGAATAGCT ATGGGAAGAACA
AVPEX2.154	AACTACCTGC(C/T)GTCGCCCTGC	TGCCAGGAGGAGAACTACC T	GGACTGGCAGGGCGA	TATCATTCTGTGGTCG GCGCCCAGGAGGAGA ACTACCTGC
AVPR2EX2.129	GCGGAGCTGG(C/T)GCTGCTCTCC	CTAGCCCGGGCGGAG	CCACAAAGACTATGGAGAG CAG	GTGGATCTTGATGTA ATGCCTAGCCCGGGC GGAGCTGG
AVPR2EX2.444	CCCATGCTGG(C/T)GTACCGCCAT	TGCCGTCCCATGCTG	CCACTTCCATGGCGGTA	GCCGTCAATGGGTGC TCAATATCTGCCGTCC CATGCTGG
BIR.1521	GGCACTTTGA(C/T)GGTGTTGCCA	AGTGGTGTGGGCACTTTG	CTACTCCAAGTTTGGCAACA C	GCCAGTCATTCCACG TATATAGAGTGGTGT GGGCACTTTGA
BIR.2463	TACCTGGGCT(T/C)GGCAGGGTCC	GGCACGGTACCTGGGC	CGCCTGGCAGAGGACC	GCCAGCCATGTGTCG AATGAGGGCACGGTA CCTGGGCT
BRS3EX1.730	CATCTATATT(A/C)CTTATGCTGT	GAAGCATTGTGTGCCATCTA	CCACTGAAATGATCACAGC A	ATCTCAGAGTGGCAT CGGATAGAAGCATTG TGTGCCATCTATATT
CAL/CGRPEX4.30	TTCCCTGCAG(C/A)CTGGACAGCC	CTGGTATGTGTTTTCCCTGC	CTTAGATCTGGGGCTGTCC	GTCTGCAATTATCGG CTGTGTCTGGTATGT GTTTTCCCTGCAG
CaR.AA1011	GACCCGACAC(C/G)AGCCATTACT	CGATACGCTGACCCGACA	GCAGCGGGAGTAATGGC	GGTCTGCATTCGCTG ATATGAGCGATACGC TGACCCGACAC
CaR.AA990	CATGGCCCAC(G/A)GGAATTCTAC	CTTTGATGAGCCTCAGAAG AA	GGAGTTCTGGTGCGTAGAAT TC	GCGAATTGAAGCCAG TTGCAAGAAGAACGC CATGGCCCAC

CHYEX2.168	ACGGCTGCTC(A/G)TTGTGCAGGA	TGTGCTGACGGCTGCT	TGTCTCACCTTCCTGCACA	CCATCGAATCGTCTA TCAGTACTTTGTGCTG ACGGCTGCTC
CLCNKBEX10.33	GGCCACCTTG(G/C)TTCTCGCCTC	CCGCTCTGGCCACCTT	AGGTGATGGAGGCGAGA	GGTCTCAATTAGGCT TCATGTACTCCGCTCT GGCCACCTTG
CLCNKBEX15.64	GCCAAGGACA(C/T)GCCACTGGAG	CCACACTGGCCAAGGA	CCTTGACCACCTCCTCCA	GCCGGTCATGTGCTC TGATATCACCACACT GGCCAAGGACA
CLCNKBEX4.19	AATCCCGGAG(G/C)TGAAGACCAT	GGTTCTGGAATCCCGGA	CCGCCAACATGGTCTTC	GCGTGATATTCCATG ATCTGAGGTTCTGGA ATCCCGGAG
CLCNKBEX4.70	GGATATCAAG(A/C)ACTTTGGGGC	TGGAGGACTACCTGGATAT CAA	CCACTTTGGCCCCAAA	GCTGGTGATGGCTCT TCATATGGAGGACTA CCTGGATATCAAG
COX2EX1.358	CCAATTGTCA(T/G)ACGACTTGCA	CGGTTAGCGACCAATTGTC	GACGCTCACTGCAAGTCG	CGAACATCTGTCACA ATGCGCTCGGTTAGC GACCAATTGTCA
COX2EX10.156	ATGGTAGAAG(T/C)TGGAGCACCA	TTTGGTGAAACCATGGTAG AA	TCAAGGAGAATGGTGCTCC	GACTCTAGTGTCGTCT GATCTCTTTGGTGAA ACCATGGTAGAAG
CYP11BIEX4.205	AGGAGCACTT(T/G)GAGGCCTGGG	AAGGTGTGGAAGGAGCACT	ATGCAGTCCCAGGCCT	TCAGATGTTGTAATC GTGCGCAAGGTGTGG AAGGAGCACTT
CYP11BIEX5.107	CGTGGCGGAG(C/G)TCCTGTTGAA	CAGTACACCAGCATCGTGG	AGTTCCGCATTCAACAGG	GCGTCGGCTTCATGC GATATTACACCAGCA TCGTGGCGGAG
CYP11B2EX3.152	CAGGCCCTGA(A/G)GAAGAAGGTG	GCAGTGGCCAGGGACT	CGTTCTGCAGCACCTTCTT	ATGCACGATCCTCTA CATTGGGACTTCTCCC AGGCCCTGA

CYP11B2EX6.91	GTGCAGCAGA(T/C)CCTGCGCCAG	CCCGACGTGCAGCAG	GCTCTCCTGGCCGCAG	CTTACCCATGATTAG CGCAGGGAACCCCGA CGTGCAGCAGA
CYP11B2EX7.65	GAGCGAGTGG(T/C)GAGCTCAGAC	GCTCTACCCTGTGGGTCTGT	TGAAGCACCAAGTCTGAGC T	GCCGATGGTGCGTCT ACTATGTCTGTTTTTG GAGCGAGTGG
DBHEX3.153	GCCCTCAGAC(G/A)CGTGCACCAT	CCGGAGTTGCCCTCAGA	GGACCTCCATGGTGCAC	TGGCAGGTTGTGACT CTCTCAACCGGAGTT GCCCTCAGAC
DBHEX4.132	GATGAAACCC(G/A)ACCGCCTCAA	GCGACTCCAAGATGAAACC	GGCAGTAGTTGAGGCGG	TATGATTATTGAGTG CGGCCCTGCGACTCC AAGATGAAACCC
DBHEX5.39	AGCCGGCCTT(G/T)CCTTCGGGGG	CCAGAGGAAGCCGGC	CCTGGACCCCCGAAG	TCAGATCGTCTTGCTG TCGAACCCAGAGGAA GCCGGCCTT
DD1R.122	CTCAGAGGAC(A/G)ACCCCGAGTA	TGACCCCTATTCCCTGCT	CTCTGACACCCCTCAAGTTC	TTTGAGATTTGTCGA GAGCCACTGACCCCT ATTCCCCTGCTT
EDNRBEX3.144	GATATAATTA(C/T)GATGGACTAC	CTGAAGCCATAGGTTTTGAT ATAAT	GCAGATAACTTCCTTTGTAG TCCA	GCCTGCTGTGGCTGT ATATCAGATAACTTC CTTTGTAGTCCATC
ELAM1.77	GACTTTCTGC(C/T)GCTGGACTCT	CCTTGGTAGCTGGACTTTCT G	GTCAGGAGGGAGAGTCCAG	GATCACTGTGGTCCC TGTCTGTAGCTGGAC TTTCTGC
ELAM1EX5.197	TTGGGACAAC(G/A)AGAAGCCAAC	CATCTGGGAATTGGGACAA	TCTACCTTTACACGTTGGCT TC	TATGAGTGTTGCGCT ATGCCTCATCTGGGA ATTGGGACAAC
ELAM1EX7.200	GTGGGACAAC(G/C)AGAAGCCCAC	CACAGGGGAGTGGGACA	CCTTCACATGTGGGCTTC	GCGTCGCTGTCGTGT ACTATCCACAGGGGA GTGGGACAAC

eNOS.78	CCCCAGATGA(T/G)CCCCCAGAAC	TGCAGGCCCCAGATG	CAGAAGGAAGAGTTCTGGG G	ATACGGGATGATGAG CATACTGCTGCAGGC CCCAGATGA
ETIEX5.90	TGAAAGGCAA(G/T)CCCTCCAGAG	TCCCAAGCTGAAAGGCA	CACATAACGCTCTCTGGAGG	TACATGACTTGCCCT GCTGTTTCATGATCCC AAGCTGAAAGGCAA
GALNREX1.327	GCACGCAGCC(G/C)CTCCGGGAGC	CAGGTGCAGCACGCA	TCCCTGGCTCCCGGA	ACGATGAGCAGGGAT CACTAACAGGTGCAG CACGCAGCC
GALNREX1.553	TCAGAAGGTC(G/C)CGGCGCAAAG	CCCACCCTCTCTCAGAAGGT	CACCGTCTTTGCGCC	ATCTGAGAGCTAGTC GGCATCCACCCTCTCT CAGAAGGTC
GGREX9.29	AACATGGGCT(T/G)CTGGTGGATC	AGCAATGACAACATGGGC	CCGCAGGATCCACCA	GGTGACTATTCGGCT GCTCTACCAGCAATG ACAACATGGGCT
GLUT2EX1.137	AGCACTAATT(C/A)TCTGTGGAGC	CTAAACAGAAACACCACAG CAC	ACTGCACTCTGCTCCACAG	TAGCTGTGTTGACAT CTGGCACAGAAACAC CACAGCACTAATT
GLUT2EX1.164	CAGTGTGCCT(T/C)CCATGCTCCA	GCAGAGTGCAGTGTGCC	GCTGTGCTGTGGAGCATG	TGCTTAGTTGTGAGT CGCCAGAGCAGAGTG CAGTGTGCCT
GLUT4EX3.112	GGCACCCTCA(C/G)CACCCTCTGG	CCCTCCAGGCACCCTC	AGAGGGCCCAGAGGGT	CTCACGACTGGGCTG ATGATTCCATCCCTCC AGGCACCCTCA
GMP-140.105	TTTCTCTTGT(A/G)ACAATGGCTT	TGGAGCGGTGGCTTCTA	CCCACCCATTATCAGACCTA	TGGCACAGTTTCCTG CTGGTGGCTCCACCT GTCATTTCTCTTGT
GMP-140.164	CCACTGGTCA(A/C)CTACCGTGCC	AAGAGAATGGCCACTGGTC	GCAGGTTGGCACGGTA	GCTGGGTGTGATCCT CTCTACAAGAGAATG GCCACTGGTCA

GMP-140.25	TACTCCAGGG(T/G)TGCAATGTCC	GCATCACTTCCTACTCCAGG	TGAGGGCTGGACATTGC	GGTGACAGTGTATTA TCTGCATCACTTCCTA CTCCAGGG
GMP-140.30	GAAGCCCCCA(G/A)TGAAGGAACC	CAGCACCTGGAAGCCC	ACACAGTCCATGGTTCCTTC	GATCTGTTCAAAGTG ATGGCGTCAGCACCT GGAAGCCCCCA
GNB3EX10.144	CATCATCTGC(G/A)GCATCACGTC	CCCACGAGAGCATCATCTG	CACTGAGGGAGAAGGCCA	TATCTTATTCTCGACG CGGCTCCCACGAGAG CATCATCTGC
GSY1EX16.210	CATCCGTGCA(C/G)CAGAGTGGCC	GCGCAACATCCGTGC	TGCAGGACGCTCGGC	CCTGTCTACCATGCA GTAATCGGCGCAACA TCCGTGCA
GSY1EX3.117	GGAGCGCTGG(A/G)AGGGAGAGCT	GGCCCTGGAGCGCTG	GGTATCCCAGAGCTCTCCC	TATATGCAGTGGTGT TCGCCTATCCCAGAG CTCTCCCT
GSY1EX8.43	AACGGCAGCG(A/G)GCAGACAGTG	GCAGGTGAACGGCAGC	AAGAAGGCAACCACTGTCT G	GACGCGGGTGCTCAT CATATCTGCGCAGGT GAACGGCAGCG
HUMGUANCYC.2 905	ATGTTGGGCT(C/G)AGGACCCAGC	TGGAGCGATGTTGGGC	CGCTCAGCTGGGTCC	GCTGGGCATGTGTAC TACTCTGATGGAGCG ATGTTGGGCT
ICAM1.117	TTCCCTGGAC(G/A)GGCTGTTCCC	CGTGGTCTGTTCCCTGGA	CCTCCGAGACTGGGAACA	CTGTCAATGCGTCTG CTCTAGACCGTGGTC TGTTCCCTGGAC
ICAM1EX1.683	TTGCAACCTC(A/C)GCCTCGCTAT	TGCTACTCAGAGTTGCAACC T	GGGAGCCATAGCGAGG	GTCTCGTCTTCGTGAG TGCAGCTACTCAGAG TTGCAACCTC
ICAM1EX2.115	GACCAGCCCA(A/T)GTTGTTGGGC	CACCTCCTGTGACCAGCC	GGGTCTCTATGCCCAACAA	ACGCACACTGATAAC TATGCACCTCCTGTG ACCAGCCCA

ICAM1EX6.254	GGTCACCCGC(G/A)AGGTGACCGT	AGGGGAGGTCACCCG	AGCACATTCACGGTCACC	GTGCTGGGTTCGCAT TCATCGCACATTCAC GGTCACCT
ICAM1EX6.39	GATGGCCCCC(G/A)ACTGGACGAG	TTTTCCAGATGGCCCC	GACAATCCCTCTCGTCCAG	CCAATAGGTGCTCAC GTCATGTGTTTTTCCA GATGGCCCCC
ICAM2EX2.63	AAAGAAGCTG(G/A)CGGTTGAGCC	CGTGAGGCCAAAGAAGCT	CCCTTTGGGCTCAACC	TTGGCTCATTTGCATG GCGCCACGTGAGGCC AAAGAAGCTG
IRS-2.AA1057	CAGGGCCCGG(G/A)CGCCGCCTCA	GTTGCCACCGCCCAG	CAACGATGAGGCGGC	TGCTCGCTTGTGATCG ACTGTTGCCACCGCC CAGGGCCCGG

KLKEX3.253	GTTGCCCACC(C/G)AGGAACCCGA	TCGTGGAGTTGCCCAC	CCCCACTTCGGGTTCC	CCTGTCGCGCCTGAT AGAATGTCGTGGAGT TGCCCACC
KLKEX4.110	GTCCAGAAGG(T/A)GACAGACTTC	AAGCCCACGTCCAGAAG	CGACACACAGCATGAAGTC TG	ACGCAATATCGGCCA TCGTGGCAAAAAAGC CCACGTCCAGAAGG
LAM1.103	CCAGTGTCAG(T/C)TTGGTAAGTC	TGGGCCCCAGTGTCA	GCAAAGAAAGGAAAGAGAC TTACC	CTGTGCCCTGCTCTGA TGATTACTATGGGCC CCAGTGTCAG
LPL.177	TATGAGATCA(A/G)TAAAGTCAGA	CAACAATCTGGGCTATGAG ATC	TGCTTCTTTTGGCTCTGACTT	GTGCCTGTTGACATA TAGTGACAATCTGGG CTATGAGATCA
LPL.98	AATAAGAAGT(C/G)AGGCTGGTGA	CCATGACAAGTCTCTGAATA AGAA	CCAGAATGCTCACCAGCC	CCTGTAGTGCAGTCT CCTGACGCATGACAA GTCTCTGAATAAGAA GT
mACEEX13 R.NA	TGCTGGTCCC(C/T)AGCCAGGAGG	GCACCCTCTGCTGGTCC	TGACTGTCACCTGTTGGGA	CACTCACTGGCACGG TATAGTGTTGGGATG CCTCCTGGCT
MRLEX2.545	CATGCGCGCC(A/G)TTGTTAAAAG	GGTGGCGTCATGCGC	CACATGATAGGGCTTTTAAC AAT	GGAATGTCTGCCGTG CCATAATGGTGGCGT CATGCGCGCC
NAT2.346	CAGGTGACCA(T/C)TGACGGCAGG	CCTTCTCCTGCAGGTGACC	ACAATGTAATTCCTGCCGTC	CTGTGAGTGATGTAC GCTCCTTCTCCTGCAG GTGACCA
NETEX11.123	AGTCCTGCCT(T/G)CCTCCTGGTG	GAAGTTCGTCAGTCCTGCC	CCCTGCAGACACTACACACC	GCGTGCGGTTCATCT GCATTCTGGAAGTTC GTCAGTCCTGCCT
NETEX12.81	CTACGACGAC(T/C)ACATCTTCCC	GCCACTCACCTACGACGA	CCAGGGCGGGAAGATG	CGGCTGGGTAGCATC ATCTAAAGCCACTCA CCTACGACGAC

NETEX5.121	AATGGCATCA(A/C)TGCCTACCTG	GAGCCTCCAATGGCATC	GTCGATGTGCAGGTAGGC	GCATGAAGTTCCATA ATCGCGAGCCTCCAA TGGCATCA
NETEX7.112	TGGTTACATG(G/C)CCCATGAACA	TCTTCTCCATCCTTGGTTAC A	TGTTGACCTTGTGTTCATGG	CAGTGACATGCCGCT CAGTACATCTTCTCCA TCCTTGGTTACATG
NETEX7.131	CACAAGGTCA(A/G)CATTGAGGAT	GCCCATGAACACACAAGGTC	TGTGGCCACATCCTCAAT	CGGCAATATGATGAT AGGTCCCCATGAACA CAAGGTCA
NETEX7.73	CACCAGCTTC(G/C)TCTCTGGGTT	GCATCAACTGTATCACCAGC TT	AGATGGCGAACCCAGAG	CCTGGTATGACATGG AGCCTCAGCATCAAC TGTATCACCAGCTTC
NETEX9.157	TGCATAACCA(A/G)GGTGAGTAGG	CGCCCTGTTCTGCATAACC	GCCCAGCCCCTACTCAC	CCAACGATGCTACTG AGTCACGCCCTGTTCT GCATAACCA
OB.160	GATCAATGAC(A/G)TTTCACACAC	AATTGTCACCAGGATCAAT GA	ACTCTCCTTACCGTGTGTGA A	CATTGCACCCACTGA GATGGATTGTCACCA GGATCAATGAC
OB-R.174	GTAATTTTCC(A/G)GTCACCTCTA	TCACATCTGGTGGAGTAATT TTC	GCTGAACTGACATTAGAGG TGA	CACGGATCTGCCGCT AGAATCATCTGGTGG AGTAATTTTCC
PG1SEX1.396	GGGAGCAGGG(T/G)TTCTCCCAGA	GCTGCGGGGAGCAGG	GGGCGCTCTGGGAGA	CGAACACATGCGGCT GGATAAGCTGCGGGG AGCAGGG
PLA2AEX2.42	GCCGCCGCCG(A/C)CAGCGGCATC	CTTGCAGTGGCCGCC	AGGGCTGATGCCGCT	AGATAGAGTCGATGC CAGCTTTGCAGTGGC CGCCGCCG
PLA2AEX3.104	TGCTGGACAA(C/A)CCGTACACCC	TGGACAGCTGTAAATTTCTG CT	ATGAATAGGTGTGGGTGTA CG	TGCCTCATTGTGACTC ATGGACAGCTGTAAA TTTCTGCTGGACAA

PNMTEX3.181	GGAGGCTGTG(A/T)GCCCAGATCT	CCTTCTGCTTGGAGGCTGT	AGCTGGCAAGATCTGGG	TGTGAGCTTGTTACT ACGGCTGCCTTCTGCT TGGAGGCTGTG	
PNMTEX3.251	GGGGGGCACC(T/A)CCTCCTCATC	GCTGAGGCCTGGGGG	CCAGGTACCACGACTCCTC	TGTGAATATGTGTGT GCCACTGAGGCCTGG GGGGCACC	
PNMTEX3.269	ATCGGGGCCCC(T/A)GGAGGAGTCG	GCTGAGGCCTGGGGG	CCAGGTACCACGACTCCTC	TGAGACTATTTAGGC TGTGCTCCTCCTCATC GGGGCCC	
PON1.584	CCCTACTTAC(A/G)ATCCTGGGAG	TCACTATTTTCTTGACCCCT ACTT	CCCAAATACATCTCCCAGGA	GATCGCAGTTCAGAG CGCATATTTTCTTGAC CCCTACTTAC	
PON2.949	AACATTCTAT(G/C)TGAGAAGCCT	CCGCATCCAGAACATTCTA	CATAAACTGTAGTCACTGTA GGCTTCT	CAGTCTCGTGGATAG CACTCGTTCTCCGCAT CCAGAACATTCTAT	
SCNN1B.222	CACCAACTTT(G/A)GCTTCCAGCC	GGAGGCCCACACCAACTT	CCGTGTCAGGCTGGAAG	GACTGGGATTACATG CTATGGAGGCCCACA CCAACTTT	
SCNN1B.238	CAGCCTGACA(C/T)GGCCCCCCCGC	TTGGCTTCCAGCCTGAC	GTGTTGGGGCTGCGG	CACTCCGATGGCGAG ATGAATTTGGCTTCC AGCCTGACA	
SCNN1B.AA442	ACCTGCATTG(G/T)CATGTGCAAG	CGCAGAGAGAGACCTGCAT T	GCAGGACTCCTTGCACAT	GCACCGTCTGTCGAT CTATACAGAGAGAGA CCTGCATTG	
SCNN1G.172	GGCTCCCCCA(T/C)GTCCAGAAGC	GGAAACAGGCTCCCCC	ACGGGGAGCTTCTGGA	AGCCAAGTGCAGGCG TACATCCTGGAAACA GGCTCCCCCA	
SCNN1G.21	TTCTGTCCAA(C/A)TTCGGTGGCC	CAGTATTGAGATGCTTCTGT CCA	CCAGCTGGCCACCGA	TCCTCTCGTTGGATGT GAGCCAGTATTGAGA TGCTTCTGTCCAA	

SCNN1GEX1.236	GTCGTGGCCC(G/T)CTCCGGGCGG	CGTTGTGAAGTCGTGGCC	CTGAGACCGCCCGGA	CAGTGACGTGAGTGC CATCTGTTGTGAAGT CGTGGCCC
SCNN1GEX2.219	GGTGTCCCGC(G/T)GCCGTCTGCG	GCATCGTGGTGTCCCG	GGAGGCGGCGCAGAC	CTCAGCAGTTAGCAG CGCATCGCATCGTGG TGTCCCGC
SCNN1GEX3.259	GCGGAAAGTC(G/A)GCGGTAGCAT	GGGAGGAAGCGGAAAGT	GAAGCCTTGTGAATGATGCT	CTTATGGCGCTGTCG GCTATCAGGGAGGAA GCGGAAAGTC
TBXASEX11.88	CCCCGCAGGC(G/A)CTGTGCTAGA	CGAGGTGCTGGGGCA	ACGGCCATCTCTAGCACA	GATATGCGTTACGTG AGTCTCGGCCATCTCT AGCACAG
TBXASEX9.276	TGCCACCTAC(C/G)TACTGGCCAC	CACACTTTCTTTTGCCACCT	AGGGTTGGTGGCCAGT	CAACAACTGCGCGAC GATGAAACACACTTT CTTTTGCCACCTAC
TGF-B1.75	CTCATGGCCA(C/T)CCCGCTGGAG	TCCTGCTTCTCATGGCC	GGCCCTCTCCAGCGG	TTGTGCATTGTTGGA CGCCCCTTTCCTGCTT CTCATGGCCA
TRHREX1.56	GCAGAACTTA(G/C)ATGATAAGCA	CAGGTACTAGAGTTTCTGCA GAACTT	GGCTTTGTCGTTGCTTATCA	AGCAGTAATGACAGC GTGCAAGGTACTAGA GTTTCTGCAGAACTT A

References:

1. Singer-Sam & Riggs, *Methods Enzymol. 225*:344-351 (1993).
2. Greenwood & Burke, *Genome Res. 6*:336-348 (1996).
3. Halushka *et al.*. Pattern of single nucleotide polymorphisms in human genes. *Nature Genet.* (Submitted).
4. Risch & Merikangas, *Science 273*:1516-1517 (1996).
5. Collins *et al.*, *Science 278*:1580-1581 (1997).
6. Collins *et al.*, *Science 282*:682-689 (1998).
7. Chakravarti, *Nature Genet. 21*:56-60 (1999).
8. Chee *et al.*, *Science 274*:610-614 (1996).
9. Wang *et al.*, *Science 280*:1077-1082 (1998); http://wwwgenome.wi.mit.edu/SNP/human/index.html.
10. Lipshutz *et al.*, *Nature Genet. 21*:20-24 (1999).
11. Shoemaker *et al.*, *Nature Genet. 14*:450-456 (1996).
12. Giaever *et al.*, *Nature Genet. 21*:278-283 (1999).
13. Pastinen *et al.*, *Clin. Chem. 42*:1391-1397 (1996).
14. Pastinen *et al.*, *Genome Res. 7*:606-614 (1997).
15. Pastinen *et al.*, *Hum. Mol. Genet. 7*:1453-1462 (1998).
16. Nikiforov *et al.*, *PCR Methods and Applications 3*:285-291 (1994).
17. Nikiforov *et al.*, *Nucleic Acids Res. 22*:4167-4175 (1994).
18. Head *et al.*, *Nucleic Acids Res. 25*:5065-5071 (1997).
19. Tobe *et al.*, *Nucleic Acids Res. 24*:3728-3732 (1996).
20. Delahunty *et al.*, *Am. J. Hum. Genet. 58*:1239-1246 (1996).
21. Chen *et al.*, *Genome Res. 8*:549-556 (1998).
22. Lyamichev *et al.*, *Nature Biotech. 17*:292-296 (1999).
23. Newton *et al.*, *Nucleic Acids Res. 17*:2503-2516 (1989).
24. Lo *et al.*, *Nucleic Acids Res.* 19:3561-3567 (1991).
25. Zhang *et al.*, *Proc. Natl. Acad. Sci. USA 89*:5847-5851 (1992).

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

CLAIMS

What is claimed is:

1. An oligonucleotide array comprising one or more oligonucleotide tags fixed to a solid substrate, wherein each oligonucleotide tag comprises a unique known arbitrary nucleotide sequence of sufficient length to hybridize to a locus-specific tagged oligonucleotide, wherein the locus-specific tagged oligonucleotide has at its first end nucleotide sequence which hybridizes to, e.g., is complementary to, the arbitrary sequence of the oligonucleotide tag.

2. A kit comprising:

(a) an array comprising one or more oligonucleotide tags fixed to a solid substrate, wherein each oligonucleotide tag comprises a unique known arbitrary nucleotide sequence of sufficient length to hybridize to a locus-specific tagged oligonucleotide; and

(b) one or more locus-specific tagged oligonucleotides, wherein each locus-specific tagged oligonucleotide has at its first (5') end nucleotide sequence which hybridizes to, e.g., is complementary to, the arbitrary sequence of a corresponding oligonucleotide tag on the array, and has at it's second (3') end nucleotide sequence complementary to target polynucleotide sequence in a sample.

3. A method of genotyping a nucleic acid sample at one or more loci, comprising the steps of:

(a) obtaining a nucleic acid sample to be tested;

(b) combining the nucleic acid sample with one or more locus-specific tagged oligonucleotides under conditions suitable for hybridization of the nucleic acid sample to one or more locus-specific tagged

oligonucleotides, wherein each locus-specific tagged oligonucleotide comprises a nucleotide sequence capable of hybridizing to a complementary sequence in an oligonucleotide tag and a nucleotide sequence complementary to the nucleotide sequence 5' of a nucleotide to be queried in the sample, thereby creating an amplification product-locus-specific tagged oligonucleotide complex;

(c) subjecting the complex to a single base extension reaction, wherein the reaction results in the addition of a labeled ddNTP to the locus-specific tagged oligonucleotide, and wherein each type of ddNTP has a label that can be distinguished from the label of the other three types of ddNTPs;

(d) contacting the complex with an oligonucleotide array comprising one or more oligonucleotide tags fixed to a solid substrate under suitable hybridization conditions, wherein each oligonucleotide tag comprises a unique arbitrary sequence complementary and of sufficient length to hybridize to a complementarysequence in a locus-specific tagged oligonucleotide, whereby the complex hybridizes to a specific oligonucleotide tag on the array; and assaying the array to determine the labeled ddNTPs present in the complex hybridized to one or more oligonucleotide tags,

thereby determining the genotype of the queried nucleotide in the sample.

4. A method to aid in determining a ratio of alleles at a polymorphic locus in a sample, comprising the steps of:

(a) using a pair of primers to amplify a region of a nucleic acid in a sample, wherein the region comprises a polymorphic locus, whereby an amplified DNA product is formed;

(b) labeling an extension primer by a single base extension reaction to form a labeled extension primer, wherein the amplified DNA product is used as a template, wherein the extension primer comprises a 3' portion and a

5' portion, wherein the 3' portion is complementary to the amplified DNA product and terminates one nucleotide 5' to the polymorphic locus, wherein the 5' portion is not complementary to the amplified DNA product, whereby a labeled dideoxynucleotide which is complementary to the polymorphic locus is coupled to the 3' end of the extension primer, wherein each type of dideoxynucleotide present in the reaction bears a distinct label; and

(c) hybridizing the 5' portion of the extension primer to one or more probes complementary to the 5' portion which are immobilized to known locations on a solid support.

5. The method of claim 4 wherein two complementary strands of the amplified DNA product are present in the single base extension reaction.

6. The method of claim 4 wherein two complementary strands of the amplifed DNA product are used as templates in the step of labeling.

7. The method of claim 4 wherein the label is a fluorescent label.

8. The method of claim 4 wherein the label is a radiolabel.

9. The method of claim 4 wherein the label is an enzyme label.

10. The method of claim 4 wherein the label is an antigenic label.

11. The method of claim 4 wherein the label is an affinity binding partner.

12. The method of claim 4 further comprising the step of:

(d) optically detecting a fluorescent label on the solid support.

13. The method of claim 4 wherein the step of labeling employs at least two distinct dideoxynucleotides bearing distinct labels.

14. The method of claim 4 wherein the step of labeling employs four distinct dideoxynucleotides bearing distinct labels.

15. The method of claim 4 further comprising the steps of:

(d) comparing quantities of a first and a second label at a location on the solid support; and

(e) determining the ratio of nucleotides present at the polymorphic locus in the sample.

16. The method of claim 15 wherein the ratio of nucleotides present at two or more polymorphic loci is determined simultaneously.

17. The method of claim 4 wherein the sample comprises DNA from two or more individuals.

18. The method of claim 17 wherein the ratio of nucleotides present at two or more polymorphic loci is determined simultaneously.

19. The method of claim 4 wherein the solid support is selected from the group consisting of beads, microtiter plates, and oligonucleotide arrays.

20. A set of primers for use in determining a ratio of nucleotides present at a polymorphic locus, comprising:

(a) a pair of primers which when in the presence of a DNA polymerase amplify a region of double stranded DNA. wherein the region comprises a polymorphic locus; and

(b) an extension primer which comprises a 3' portion which is complementary to a portion of the region of double stranded DNA and a 5' portion which is not complementary to the region of double stranded DNA, wherein the extension primer terminates one nucleotide 5' to the polymorphic locus.

21. A kit comprising in a single container two or more of the sets of primers of claim 20.

22. A kit comprising in a single container:

(a) a set of primers of claim 20; and

(b) a solid support comprising a probe which is attached to a solid support, wherein the probe is complementary to the 5' portion of the extension primer.

23. The kit of claim 22 wherein the solid support is an oligonucleotide array.

24. The kit of claim 22 wherein the solid support is a bead.

25. The kit of claim 22 wherein the solid support is a microtiter plate.

26. A method to aid in determining a ratio of alleles at a polymorphic locus in a sample, comprising the steps of:

(a) labeling an extension primer by a single base extension reaction to form a labeled extension primer, using a DNA molecule as a template, wherein the extension primer comprises a 3' portion and a 5' portion, wherein the 3' portion is complementary to the DNA molecule and terminates one nucleotide 5' to a polymorphic locus, wherein the 5' portion is not complementary to the DNA molecule, whereby a labeled

dideoxynucleotide which is complementary to the polymorphic locus is coupled to the 3' end of the extension primer, wherein each type of dideoxynucleotide present in the reaction bears a distinct label; and

(b) hybridizing the 5' portion of the extension primer to one or more probes complementary to the 5' portion which are immobilized to known locations on a solid support.

27. The method of claim 26 wherein two complementary strands of the DNA molecule are present in the single base extension reaction..

28. The method of claim 27 wherein each complementary strand of the DNA molecule is used as a template to label an extension primer.

29. The method of claim 26 wherein the label is a fluorescent label.

30. The method of claim 26 wherein the label is a radiolabel.

31. The method of claim 26 wherein the label is an enzyme label.

32. The method of claim 26 wherein the label is an antigenic label.

33. The method of claim 26 wherein the label is an affinity binding partner.

34. The method of claim 26 further comprising the step of:

(c) optically detecting a fluorescent label on the solid support.

35. The method of claim 26 further comprising the steps of:

(c) comparing quantities of a first and a second label at a location on the solid support; and

(d) determining the ratio of nucleotides present at the polymorphic locus in the sample.

36. The method of claim 35 wherein the ratio of nucleotides present at two or more polymorphic loci is determined simultaneously.

37. The method of claim 26 wherein the sample comprises DNA from two or more individuals.

38. The method of claim 34 wherein the ratio of nucleotides present at two or more polymorphic loci is determined simultaneously.

39. The method of claim 26 wherein the step of labeling employs at least two distinct dideoxynucleotides bearing distinct labels.

40. The method of claim 26 wherein the step of labeling employs four distinct dideoxynucleotides bearing distinct labels.

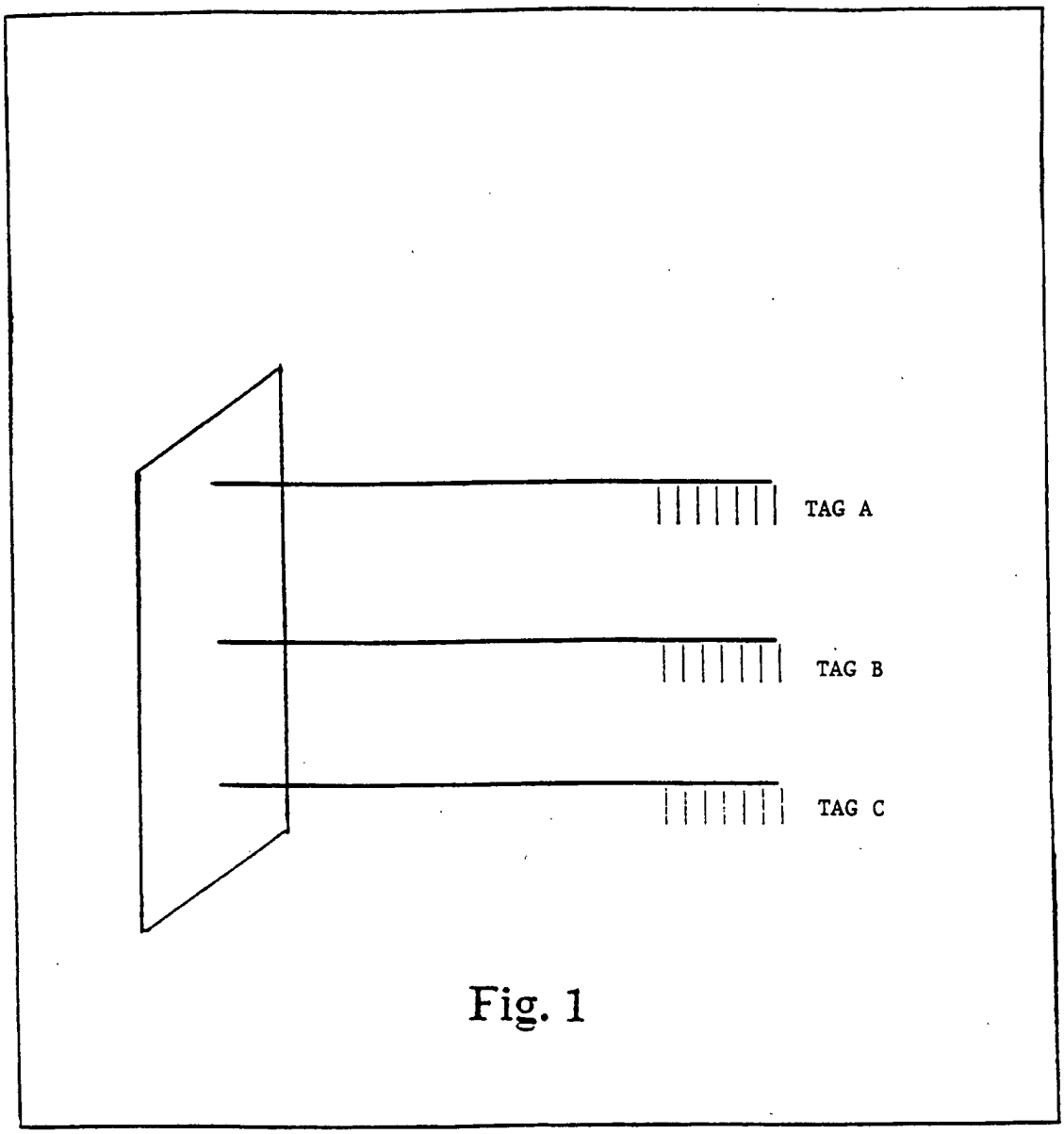


Fig. 1

Locus-specific tagged oligonucleotide:

synthetic sequence specific for a particular oligonucleotide tag on the array (e.g., "Tag A," "Tag B," "Tag C," etc.)

sequence specific for the amplification product of a particular SNP (e.g., SNP "A," SNP "B," SNP "C," etc.)

Fig. 2

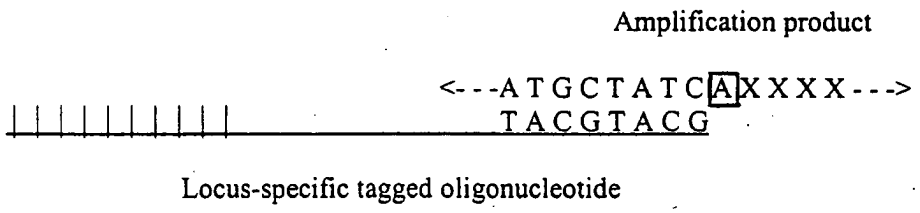

Amplification product
<- - -A T G C T A T C A X X X X - - ->
T A C G T A C G
Locus-specific tagged oligonucleotide

Fig. 3

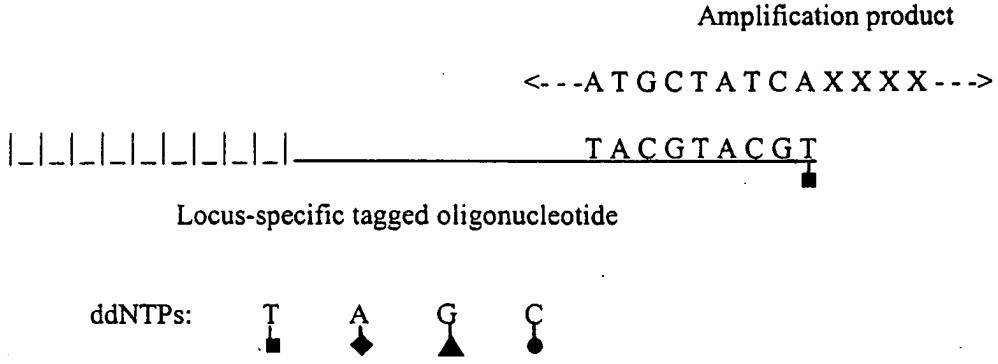

Amplification product
<- - -A T G C T A T C A X X X X - - ->
T A C G T A C G T
Locus-specific tagged oligonucleotide
ddNTPs:
T
A
G
C

Fig. 4

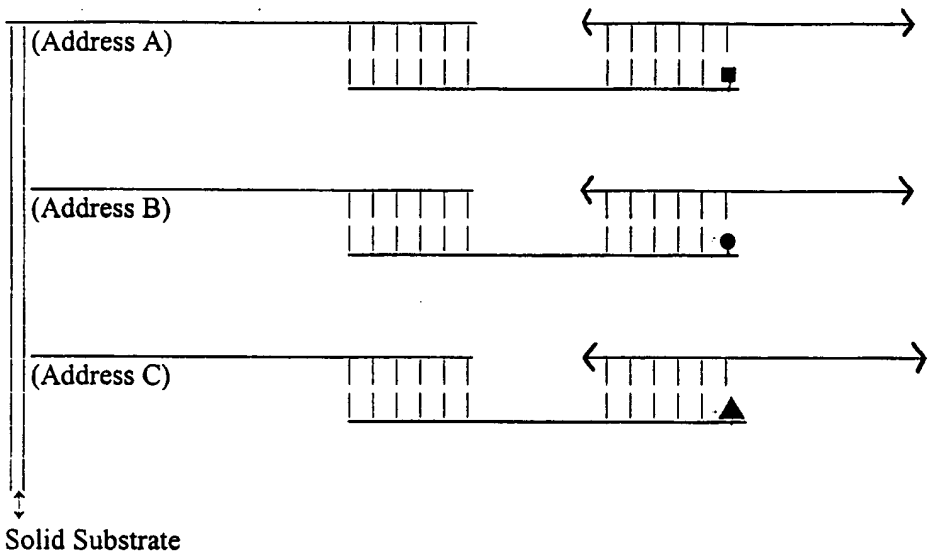


Fig. 5

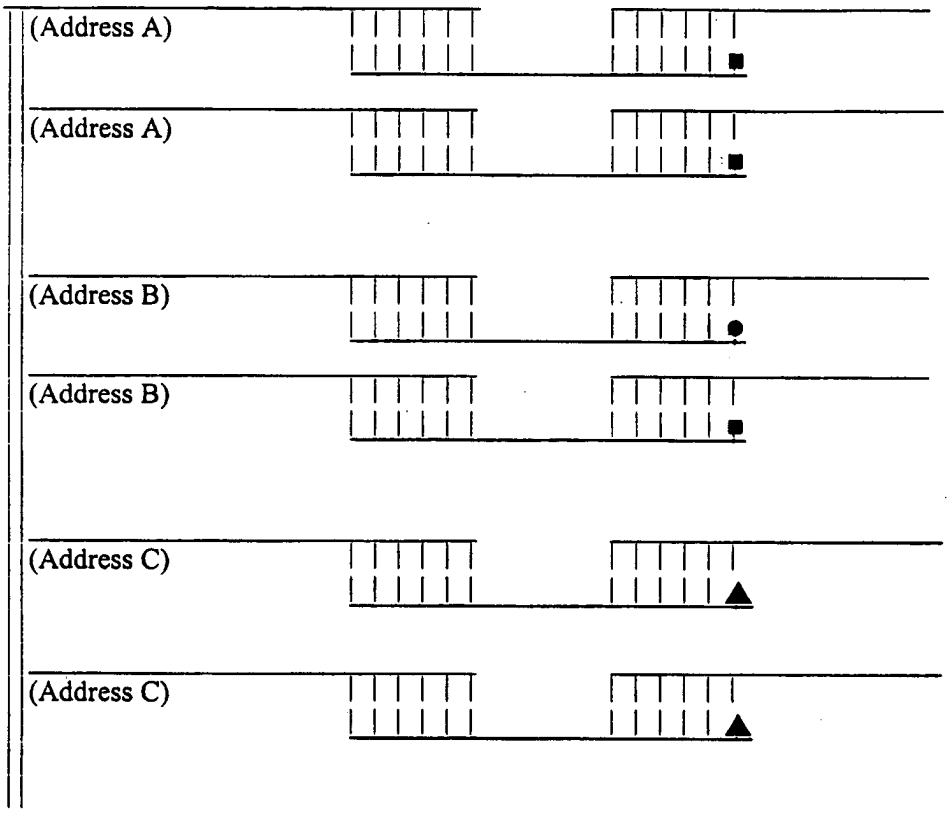


Fig. 6

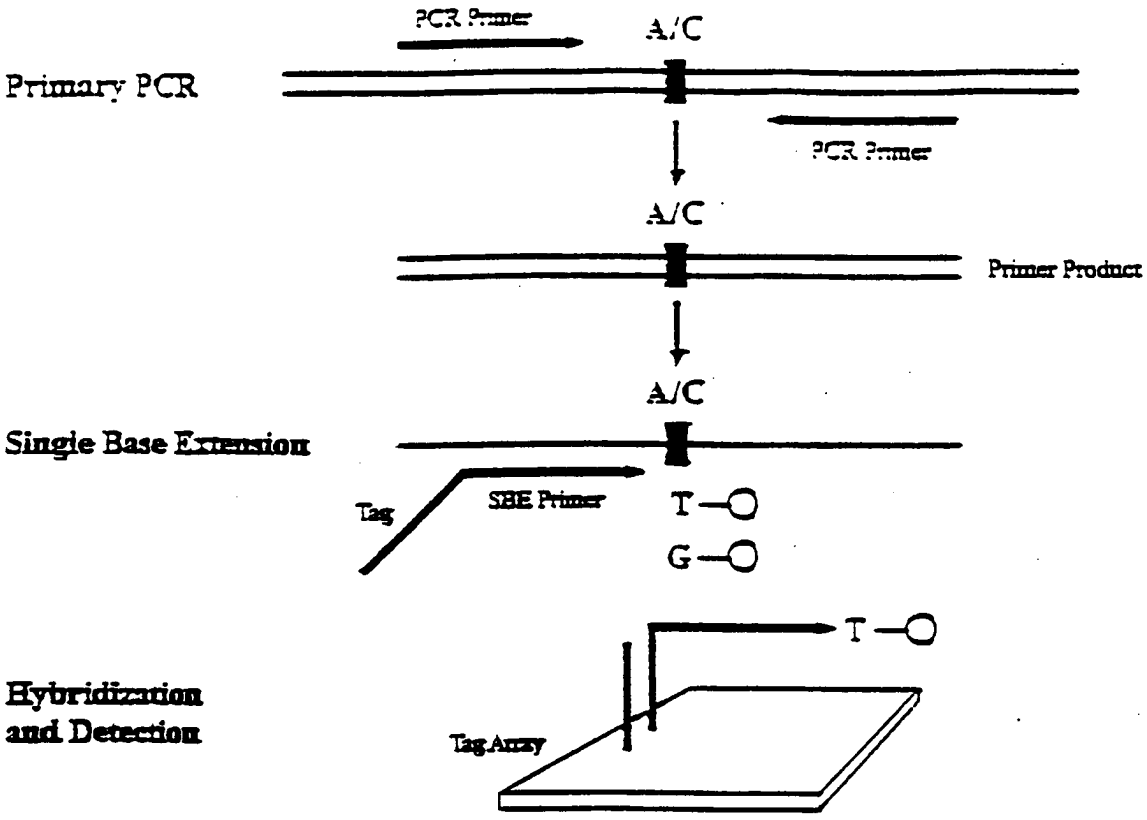

PCR Primer
A/C
Primary PCR
PCR Primer
A/C
Primer Product
A/C
Single Base Extension
Tag
SBE Primer
T
G
T
Hybridization
and Detection
Tag Array

Fig. 7

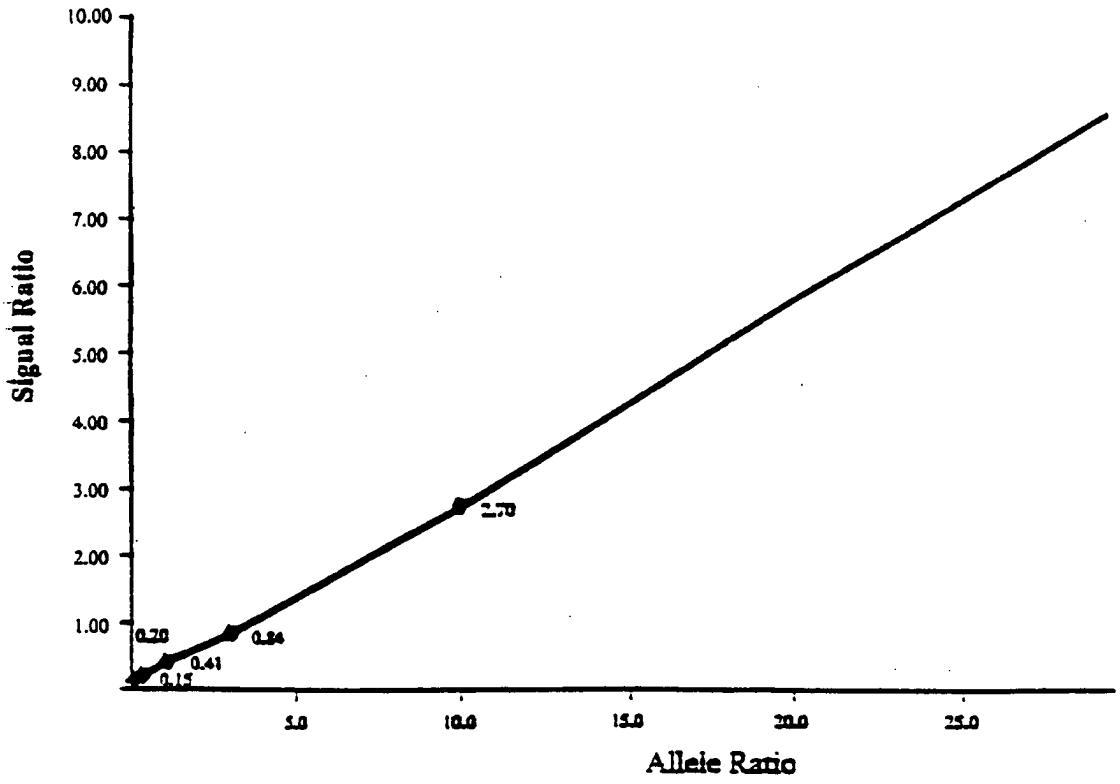

10.00
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
Signal Ratio
0.20
0.41
0.84
2.70
5.0
10.0
15.0
20.0
25.0
Allele Ratio

Fig. 8

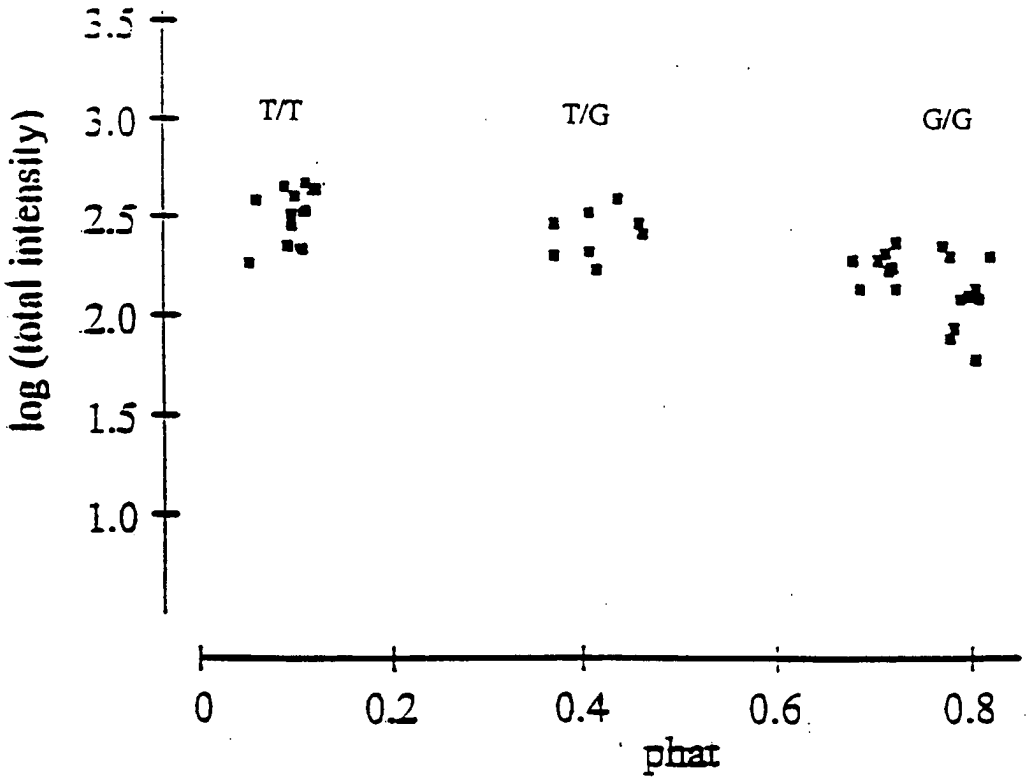


Fig. 9